

us-09-668-119-1.rni

RESULT

4536.338 Million cell updates/sec

TITLE OF INVENTION:	Gene Sequence for Spinocerebellar Ataxia
Patent No.	5741645
TITLE OF INVENTION:	Type 1 and Method for Diagnosis

CORRESPONDENCE ADDRESS:

SHEET: 119 NO. 3

STATE: MN

ZIP: 55401
COMPUTER READER

COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Patentin Release

```

APPLICATION NUMBER: US/0

FILED DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muelling, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110-00030107

```

; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 612-305-1217
;
; TELEFAX: 612-305-1225
;
; INFORMATION FOR SEQ ID NO: 3:

```

LENGTH: 234 base pairs

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

Query Match          7.5%; Score 130.8; DB 1; Length 234;
Best Local Similarity 75.7%; Pred. No. 2.8e-20;
Matches 162; Conservative 0; Mismatches 52; Indels 0; Gaps 0

QY 370 cagaccagctgaagcttccagcagatgtgtggtctgcagcagcagcagcagcaatc 429
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

QY 430 cagcagcagcagcagcagcagcggtcactacagcagcagcagcagcagcaatc 489
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

QY 490 caggtcagcagagtgccatgcagcagcagcttccaaagcagtagtgcagcagcagcag 549
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180

QY 550 ctccagcagcagcagcagcagcagcagcagcaatc 583
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 214

```

File copy

OM of: US-09-668-119-1 to: A_Geneseq_1101.* out_format: pfs
Date: Feb 28, 2002 7:22 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODL=frame+ n2p_model -DEV=xip
-O=/sgn2_1/USPRO.spool/US09668119/runat.28022002.084716.19265/app-query.fasta.1.1824
-DB=A_Geneseq_1101 -QPM=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09668119 -SCGNL_1.87 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-668-119-1
Query Length: 1740
Database: A_Geneseq_1101.*
Database sequences: 522463
Database Length: 74073290
Search time (sec): 99.080000

Score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM40299		2859.00	3024.00	3.2e-167	748
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM42085		2860.00	3024.00	1.2e-161	780
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM41262		390.00	425.49	1.2e-15	79
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM46612		390.00	425.49	1.2e-15	79
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM01992		390.00	425.49	1.2e-15	79
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT.AAM78866		389.00	401.73	2.1e-15	1004
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT.AAM52830		389.00	401.73	2.1e-15	1004
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB76123		376.50	405.51	8.5e-15	149
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB69612		353.50	380.77	2.0e-13	155
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB69614		345.00	372.35	6.2e-13	145
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB84634		342.00	341.58	1.4e-12	3190
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY4791		341.00	353.88	1.2e-12	233
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAB95124		336.50	363.64	7.4e-12	1337
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY81609		329.50	336.25	5.7e-12	121
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB69609		328.50	332.45	1.1e-11	1326
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT.AAY55933		323.50	325.81	1.7e-11	1185
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY54319		312.50	319.14	6.9e-11	1185
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT.AAY3497		312.00	313.66	8.0e-11	2074
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM40064		311.50	328.49	6.7e-11	369
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM41252		311.50	328.49	6.7e-11	369
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY54320		310.00	331.33	1.1e-10	2023
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY54320		307.50	331.33	1.1e-10	250
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY82335		306.50	326.14	2.3e-10	2783
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY82335		304.50	303.04	2.3e-10	2783
/SID22/gcgdata/geneseq/geneseq/AA1992.DAT.AAR23963		304.50	303.04	2.3e-10	1018
/SID22/gcgdata/geneseq/geneseq/AA1992.DAT.AAR23963		304.00	311.48	2.2e-10	1700
/SID22/gcgdata/geneseq/geneseq/AA1995.DAT.AAM01496		304.00	306.90	2.2e-10	783
/SID22/gcgdata/geneseq/geneseq/AA1995.DAT.AAB18144		302.00	311.70	3.1e-10	783
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT.AAM7151		301.00	310.60	3.1e-10	802
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT.AAM7151		301.00	310.43	3.2e-10	615
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT.AAM7153		300.50	312.27	4.5e-10	230
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT.AAY40099		297.00	317.34	5.7e-10	1162
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAY62058		297.00	302.88	5.7e-10	1162
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY96255		297.00	302.88	5.7e-10	1162
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT.AAY58500		295.00	321.88	5.7e-10	113
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB62331		295.00	302.16	8.7e-10	831
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAB69615		293.50	302.16	8.7e-10	903
/SID22/gcgdata/geneseq/geneseq/AA1995.DAT.AAR80168		292.50	300.36	1.0e-09	903
/SID22/gcgdata/geneseq/geneseq/AA1997.DAT.AAMW37389		292.50	300.36	1.0e-09	903

seq_name: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.AAM40299
seq_documentation_block:
ID AAM40299 standard; Protein: 748 AA.
XN AAM40299;
XX 22-Oct-2001 (first entry)
XX Human polypeptide SEQ ID NO 3444.
XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-Jul-2001.
XX 26-Dec-2000; 2000WO-US34263.
XX 21-Jan-2000; 2000US-0488725.
XX 25-Apr-2000; 2000US-0552317.
XX 09-Jul-2000; 2000US-0598042.
XX 19-Jul-2000; 2000US-0620312.
XX 03-Aug-2000; 2000US-0653450.
XX 14-Sep-2000; 2000US-0662191.
XX 19-Oct-2000; 2000US-0693036.
XX 29-Nov-2000; 2000US-0727344.
XX (HSEB) HSEB INC.
XX Tang Y, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao GA, Zhou P, Goodrich R, Dymac RC;
XX WPI, 2001-442253/47.
XX N-PSDB: AAI59455.
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX Example 6; SEQ ID NO 3444; 10078bp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and
the encoded polypeptides (AAM38642-AA42213) with neotropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemia and
C.N.S disorders.
Note: The sequence data for this patent did not form part of the printed
specification.

Sequence 748 AA;

alignment_scores:

Quality: 2959.00 Length: 580
 Ratio: 5.111 Gaps: 0
 Percent Similarity: 99.828 Percent Identity: 99.483

alignment_block:

US-09-668-119-1 x AAM40299

Align seg 1/1 to: AAM40299 from: 1 to: 748

```

1  ATGAGCAAGCTGTGTGGCAGACAGTAATCCAGCAAGATATGACAG 50
27  MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetGlu 43
51  CCATGTTTCTGAAAGCCAGACCCGGAGATACCTTTCTGTGG 100
43  HisValPheLeuLysAlaLysThrArgAspGluTrpLeuSerLeuVal 60
101 CCAGGCTCATTTATCCATTTTCAGACATTCATACAGAAATCTCAGCT 150
60  LaArgLeuIleIleHisPheArgAspIleHisAsnLysSerGlnAla 76
151 TCCGTCAGTGTCTCTATGATGACCTCCAGAGCTGACTGGCGACCTGC 200
77  SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyProAl 93
201 TGCAGGAGCCGCTGGAATGGCATGCTCCTCGGGGGCCCGGACAGTCTC 250
93  aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSer 110
251 TGGCGGGATGGGTAGCTTTGTGTCATGGAGACAGCCATGCTCTCTCA 300
110  euGlyGlyMetGlySerLeuGlyAlaMetGlyGlnProMetSerLeuSer 126
301 GGGAGGCGGCTCCTGGGACCTGGGGATGGCCCTCAGAGCTGGCTGT 350
127  GLyGlnProProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 143
351 CGTGTCTACGGCAACTCCAGACAGCCAGCTGCAGTCCAGAGGTGGCG 400
143  ValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnAlaAla 160
401 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
160  euGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 176
451 CTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
177  LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 193
501 GAGTGCATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
193  nSerAlaMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 210
551 TCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
210  euGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 226
601 AATCAGCAACAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 650
227  AsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 243
651 GGTGAGCTCCCAACAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 700
243  nLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
701 AGCAGCAGGCTTTGGAGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 750
260  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 276
751 CAGCCAGAGCCTCCGCTCCAGAGGCTGTGCCAGAGAGAGAGAGAGAT 800

```

```

277  GlnProGlnProProProSerGlnAlaLeuProGlnGlnLeuGlnGln 293
801  GCATCACACAGACACACAGCCGCGACACAGCCCGCCAGAGCTCCAG 850
293  tHisHisThrGlnHisHisGlnProProProGlnGlnGlnGlnGlnGln 310
851  TTGCTCAGAAACCAACCATCAACAATCCCGCCAGAGTCCGAGAGCCGCT 900
310  AlaGlnAsnGlnProSerGlnLeuProProGlnSerGlnThrGlnPro 326
901  TTGGTGTACAGGCGCAGAGCTCTCCCTGACAAATGTTATACCAACC 950
327  LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuTrpThrGlnPr 343
951  ACCACTGAATTTGTCCAGCTCCGATGGTGGTGGAGAGAGCCCGCAGTGC 1000
343  oProLeuLysPheValArgAlaProMetValValGlnGlnProProValG 360
1001 AGCCCGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1050
360  InProGlnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 376
1051 CAGATGTGGTCCCGGAGTCCAGGTCCAGCCAGAGAGAGAGAGAGAGAG 1100
377  GlnMetValAlaProGlyValGlnValSerGlnSerLeuProMetLe 393
1101 GTTCCTGCGCGTCACCGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1150
393  uSerSerProSerProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 410
1151 CTCCCCCGCAGCGCTCCCGCAGAGCCCGCCAGAGAGAGAGAGAGAGAG 1200
410  roProGlnProProSerProGlnProGlyGlnProSerSerGlnProAsn 426
1201 TCCACAGTCAAGCTGTGGCCCTGGCCCATCTCCAGTAGTCTCCGCGCAG 1250
427  SerAsnValSerSerGlyProAlaProSerProSerSerPheLeuProSe 443
1251 CCCCTCACCGAGCCCTCCAGAGCCAGAGAGAGAGAGAGAGAGAGAGAG 1300
443  rProSerProGlnProSerGlnSerProValThrAlaArgThrProGln 460
1301 ACTTCAGTGTCCCTCACCTGAGCTTTAAACACACACTGTGAGACCCAGC 1350
460  snPheSerValProSerProGlyProLeuAsnThrProValAsnProSer 476
1351 TCTGTCAATGAGCCAGCTGGCTCCAGCCAGAGAGAGAGAGAGAGAGAG 1400
477  SerValMetSerProAlaGlySerSerGlnAlaGlnGlnGlnGlnGlnGln 493
1401 GGACAAGCTGAGAGAGAGCTGTGGAAGTACATCGAGCCCTCGCCGCGATGA 1450
493  uAspLysLeuLysGlnLeuSerLysTrpIleGlnProLeuArgMetI 510
1451 TCAACAAGATCGACAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
510  LeAsnLysIleAspLysAsnGlnAspArgLysLysAspLeuSerLysMet 526
1501 AAGAGCTTTGTGACATTTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1550
527  LysSerLeuLeuAspIleLeuThrAspProSerLysArgCysProLeuLys 543
1551 GACCTTCAAAAGTGTGAGATGCGCCTGGAGGAACCAAGAAAGAGAGATGC 1599
543  sThrLeuGlnLysCysGlnIleAlaLeuGlnLysLeuLysAsnAspMetA 560
1600 GGTGCCATCTCCCGCAGCGCCCGGTCGACAGCAAGAGAGAGAGAG 1649
560  LaValProThrProProProProValProProThrLysGlnGlnTrp 576
1650 CTATGCGAGCGGCTCTGTGATGCGCTGCGCACACATCGCTACCTGT 1699

```

577 LeucylGlnProLeuLeuAspAlaValLeuAlaAsnIleArgSerProVa 593

1700 CTTCAACATTCCTGCTACCGACATTCGTCCAGCCATG 1739
|||||
593 lPheAsnHisSerLeuTyrArgThrPheValProAlaMet 606

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AA42085

seq_documentation_block:

ID AA42085 standard; Protein: 780 AA.

AC AA42085;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 7016.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000MO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;

XX MPI: 2001-442253/47.

XX N-PSDB; AAT61241.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 780 AA;

alignment_scores:
Quality: 2860.00 Length: 581
Ratio: 4.948 Gaps: 0
Percent Similarity: 99.484 Percent Identity: 98.967

alignment_block:
US-09-668-119-1 x AA42085

Align seg 1/1 to: AA42085 from: 1 to: 780

1 ATGAGGAAAGCTGTGTGGACACACAGTAATCCAGCAGATATGAGAG 50
|||||
58 MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetGluSe 74
|||||
51 CCATGTTTCTGGAAGCCAGACCCGGAGCAATACCTTCTCTGCTG 100
|||||
74 rHisValPheLeuLysAlaLysThrArgAspIuTyrLeuSerLeuValA 91
|||||
101 CCAGGCTCATTTCCATTTTGGAGCATTCATTAACAAGAAATTCAGCT 150
|||||
91 laArgLeuIleIleHisPheArgAspIleHisAsnLysLysSerGlnAla 107
|||||
151 TCCGTCAGTATCTATGATGACTCCAGAGCTGAGTGGGGAGCTGC 200
|||||
108 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyProAl 124
|||||
201 TGGGGAGCCGCTGGAATTTGGATGCTCTCTGCGGGGCGGAGCTGC 250
|||||
124 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSerL 141
|||||
251 TGGGGCGGATGGGTAGCTTGTGGCATGGGAGCAGCATGCTCTCA 300
|||||
141 enGlyIleMetGlySerLeuGlyAlaMetGlyGlnProMetSerLeuSer 157
|||||
301 GGGCAGCCGCTCTCTGGGACTCTGGGAGTGGGGATGGCCCTCACAGCATGGCTGT 350
|||||
158 GlyGlnProProProGlyThrSerGlyMetAlaProHisSerMetAlaVa 174
|||||
351 CGTGTCTACGGCACTCCACAGACCCAGCTGCTCCAGCAGGTGGC. G 399
|||||
174 lValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAlaAla 191
|||||
400 CTCGACAGCAGCAGCAGCAGCAGCAGC. AGTTCCAGAGCAGCAGCAGCGG 448
|||||
191 laAlaAlaAlaAlaAlaAlaThrAlaArgSerSerSerSerSerArgArg 207
|||||
449 CGCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG. CAGTTCCAGGCTCA 497
|||||
208 ArgTyrSerSerSerSerSerSerSerSerSerSerGlnPheGlnAlaGl 224
|||||
498 GCAGAGTCCCATGACAGCAGCAGTTCACAGCAGTATGTCAGCAGCAGCAGC 547
|||||
224 ngInSerAlaMetGlnGlnGlnPheGlnAlaValAlaGlnGlnGlnGlnG 241
|||||
548 AGTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAAATTCATCAT 597
|||||
241 lInLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeuIleLysLeuHis 257
|||||
598 CAAATTCAGCAGCAGTATACAGCAGCAGCAGCAGCAGCAGTCCAGGATAC 647
|||||
258 GlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAl 274
|||||
648 ACAGCTGACAGCTCCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 697
|||||
274 agInLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 291
|||||
698 AGCAGCAGCAGGCTTGGAGCCAGCAGCAGCAGCAGTTCAGCAGCAGCAGGAT 747
|||||
291 lngInGlnGlnAlaLeuGlnAlaGlnAlaGlnProProIleGlnGlnProPromet 307
|||||
748 CAGCAGCCACAGCCTCCGCTCCAGGCTGTGCCCCAGCAGCTGACAGCA 797

308 GInGInPrOGInPrOPrOPrOSerGInAlaLeuPrOGInGInLeuInGInI 324
798 GATGATCATCACACAGACACACAGCCGCCACACAGACCCCAAGCTC 847
324 mEthIsHISrGInHISHSInGInPrOPrOGInPrOGInGInPrOP 341
848 CAGTTGGCCAAACCAACCATCAACAACTCCCGCACAGTGCAGACCCG 897
341 roValAlaGInASnGInPrOSerGIIneuPrOPrOGInSerGIInrGIn 357
898 CCTTGGTGTACAGGGCGCAAGCTCTCCAGCAATGTGTATACCA 947
358 ProeuValSerGIInAlaGInAlaLeuPrOGInGInMetLeuTythrGI 374
948 ACCACCACTGAAATTTCTCCGAGCTCCGATGGTGGTGCAGACCCCGC 997
374 nPrOPrOeuuysrheValrAlrAlaPrOMetValValGInGInPrOPrOV 391
998 TGCAGCCCCAGGTGCACACACAGACAGACAGCATACAGACAGCTAGCT 1047
391 aGInPrOGInValGInGInGInGInrAlaValaGInrHrAlaGInAla 407
1048 GCCCAAGTGGTGGCTCCCGAGTCCAGGTACAGCCAGACAGCAAGCTCCCAT 1097
408 AlaGInMetValAlaPrOGInValaGInValaSerGIInSerLeuPrOMe 424
1098 GCTGTCTCTCCCTCAACCGGGCCAGAGGTGCAGACCCCGCAGTGTGATGC 1147
424 tLeuSerSerPrOSerPrOGInGInbInrAlaGInrHrPrOGInSerMeLP 441
1148 CCCCTCCCCCCAGCGCTCCCGCAGCCGCCAGCCAGCTACAGCC 1197
441 rOPrOPrOPrOGInPrOSerPrOGInPrOGInGInPrOSerGIInPrO 457
1198 AACCTCCAACTGACTGTGGCCCTGGCCCATGCCAGTAAAGTCTTGCC 1247
458 ASnSerASnValSerSerGIInPrOAlaPrOSerPrOSerSerPheluPr 474
1248 CAGCCCTCTACGGCAGCCCTCCAGAGCCCAAGTGAGCGCGCAGACCCAC 1297
474 OSerPrOSerPrOGInPrOPhEgInSerPrOVAlThrAlaGrHrPrOG 491
1298 AGAAGCTTCAAGTGTCCCTCACCGGACCTTTAAACACACTGTGAACCC 1347
491 InASnPheserValPrOSerPrOGInPrOleuASnPrHrPrOVAlASnPrO 507
1348 AGCTGTGTATGAGCCCACTGGCTCCACCCAGGTGAGGAGCAGACATYA 1397
508 SerSerValMetSerPrOAlaGInSerSerGIInAlaGInGInGInrTy 524
1398 CCGGAGCAAGCTGGAACCACTCTGCCAAGTACATCGAGCCCTCGCCGCCA 1447
524 rLeuASnPrLrLeuVrGIInLeuSerLysrTyrlLeGIInPrOleuArGApG 541
1448 TGAATCAACAAGATCGACAAGAACGAAGACAGAAAAAGACCTGATGAG 1497
541 etIlleASnLysrIlleASnPrLysrASnGInuASnArGInLysrASnLeuSerLys 557
1498 ATGAAGAGCTTCTGACATTTGTGACACACCCCTCGAAGCGGTGTCCCT 1547
558 MetLysrSerLeuLeuASnPrLLeuLeuThrASnPrOSerLysrATrGyCSPrOle 574
1548 GAAGACCTTGCAAAAGTGTGAGATGGCCCTGGAGAAATCAAGATGACA 1597
574 uLysrThrLeuGInLysrGInLrIlleAlaLeuGIInLysrLeuLysASnASpM 591
1598 TGGG.GTGGCCCACTCCCCACCGCCGCCGTGGTCCACGAGCAACAAACAGAG 1646
591 etAlaValaPrOHrPrOPrOPrOPrOVAlPrOPrOHrLysGInGIn 607
1647 TACCTGTGCACCGCTCTGTGATCGCTCCGGCAACATCGCTCAC 1696

CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 79 AA;

alignment_scores:
 Quality: 390.00 Length: 79
 Ratio: 4.937 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-668-119-1 x AAM01992 ..

Align seq 1/1 to: AAM01992 from: 1 to: 79

```

376 CAGCTGCAGCTCCAGCGGTGGCTCCAGCAGCAGCAGCAGCA 425
|||||
1 GlnLeuInLeuGlnGlnValAlaLeuGlnGlnGlnGlnGln 17
426 GTTCACAGCAGCAGCAGCGCGCTACAGCAGCAGCAGCAGC 475
|||||
17 nphGlnGlnGlnGlnGlnAlaLeuGlnGlnGlnGlnGln 34
476 AACAGCAGCGCTTCAGGCTCAGCAGAGTCCATGCAGCAGCAG 525
|||||
34 InGlnGlnGlnInphGlnAlaGlnGlnSerAlaMetGlnGlnInphGln 50
526 GCAGTAGTCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCA 575
|||||
51 AlaValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 67
576 GCATCTAATTAAATGCATCATCAAAATCAGCAACAG 612
|||||
67 nHISLeuIleLysLeuHISHisGlnAsnGlnGln 79

```

seq_name: /SID52/gcgdata/geneseq/genesep/AA1998.DAT:AAW37856
 seq_documentation_block:
 ID AAW37856 standard; Protein; 1004 AA.
 AC
 XX AAW37856;
 XX
 DT 28-AUG-1998 (first entry)
 XX
 DE Human polyhomeotic 1 (hph1) protein.
 XX
 KW Polyhomeotic 1; hph1 gene; oncogene; human; neoplasia;
 KW dysplasia; hyperplasia; diagnosis; tumour; metastasis; therapy.
 XX
 OS Homo sapiens.
 XX
 FN M09807860-A1.
 XX
 PD 26-FEB-1998.
 XX
 PF 22-AUG-1997; 97WO-US14886.
 XX
 PR 06-FEB-1997; 97US-0036939.
 PR 23-AUG-1996; 96US-0024349.
 PR 04-DEC-1996; 96US-0031569.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Randazzo F;
 XX
 DR WPI; 1998-169164/15.

DR N-PSDB; AAV29009.

XX
 XX Isolated human polyhomeotic 1 oncogene - used to develop products
 PR for diagnosis and therapy of proliferative and developmental
 PR disorders, e.g. neoplasia, dysplasia or hyperplasia
 XX
 PS Claim 1; Page 28-30; 39pp; English.

XX This polypeptide comprises a novel human polyhomeotic 1 protein
 CC encoded by the hph1 oncogene that is implicated in neoplastic
 CC disorders. Also claimed are: (1) an hph1 polypeptide consisting of
 CC at least 22 contiguous amino acids; (2) an hph1 fusion protein; (3)
 CC antibodies which specifically bind to a hph1 protein; (4) an
 CC expression construct for expressing all or a portion of hph1
 CC protein; (5) a homologously recombinant cell; (6) a method of
 CC identifying neoplastic tissue of a human, where over-expression of
 CC the hph1 gene identifies the tissue as being neoplastic; (7) a
 CC method to aid in the diagnosis or prognosis of neoplasia, where a
 CC difference in the hph1 gene, mRNA, or protein between a first and
 CC second tissues indicates neoplasia in the first tissue; (8) a method
 CC to aid in detecting a genetic predisposition to neoplasia; (9) a
 CC method of identifying a human chromosome 12; and (10) a method of
 CC inducing a cell to de-differentiate by contact with a hph1 gene or
 CC expression product. The products can also be used to inhibit hph1
 CC growth. They can be used to treat e.g. tumours, anhydric hereditary
 CC ecodermal dysplasia, congenital alveolar dysplasia, epithelial
 CC dysplasia of the cervix, fibrous dysplasia of bone, or mammary
 CC dysplasia, endometrial, adrenal, breast, prostate, or thyroid
 CC hyperplasia or pseudo-epitheliomatous hyperplasia of the skin.
 CC Even in disorders in which hph1 mutations are not implicated,
 CC down-regulation or inhibition of hph1 expression can have therapeutic
 CC application. In these disorders, decreasing hph1 expression can help
 CC to suppress tumours. Similarly, in tumours where hph1 expression is
 CC not aberrant, effecting hph1 down-regulation can suppress metastases.
 XX
 XX Sequence 1004 AA;

alignment_scores:
 Quality: 389.00 Length: 522
 Ratio: 1.379 Gaps: 20
 Percent Similarity: 54.023 Percent Identity: 28.544

alignment_block:
 US-09-668-119-1 x AAW37856 ..

Align seq 1/1 to: AAW37856 from: 1 to: 1004

```

178 CAGAGCCTGACTGGCGGAGCCTGCTGGGAGCGCGTGGATTCGATGCC 227
|||||
260 GlnSerLeuAsnLeuSerGlnAlaGlyGlyGlySerGlyAsnSerIlePr 276
228 T.....CCTGGGGCCCGGAGCAGTCTGCGCGGATGGATTA 265
|||||
276 OGlySerMetGlyProGlyGlyGlyGlnAlaHisGlyGlyGlyGly 293
266 GCTTGTGTCGATGGGAGCAGCAATGCTCTCAAGGAGCCGCTCT 315
|||||
293 InLeuProSerSerGly.....MetGlyGlyGlySerGlyProAlaGly 307
316 GGGACCTCGGGGATGGCCCTCAGACAGATGCTGTGCTCGGCAAC 365
|||||
308 GlyThrGlyValAlaGlnProLeuProAlaAlaGlnThrValThrValSe 324
366 TCACAGACCCAGCTCAGCTCAGCAGAGTGGCGCTCAGCAGCAGC 415
|||||
324 rGlnGlySerGlnThrGlyAlaGlnSerAlaAlaAlaLysLysAlaGln 341
416 AACAGCAGCGCTTCAGCAGCAGCAGCAGCGCGCTACAGCAGCAG 465
|||||
341 LaSpGlySerGlyGlnGlnAsnValGlyMetAsnLeuThrArgThrAla 357

```

```

466 CAGCAGCAGCAGCAGCAGTTCAGGCTCAGCAGATGCTCCATGCA 515
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
358 ThrProAlaProSerGlnThrLeuIleSerSerAlaThrGlyThrGlnI 374
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
516 GCAGTTCAGCAGTACTGACAGCAGCAGCAGCAGTCCACAGCAGCAG 565
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 eGlnProHisSerLeuIleGlnGlnGlnGlnGlnIleHisLeuGlnI 391
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
566 AGCAGCAG.....CAGCATCTAATTAAATTCATCATCAAAATCAGCA 609
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
391 ysglnValValIleGlnGlnGlnIleAlaIleHisHis.....GlnGln 405
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
610 CAGATACAGCAGCAGCAGCAGCTCCAGCAATGCA...CAGCTGCA 656
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
406 GlnPheGlnHisArgIleSerGlnLeuIleHisThrAlaThrHisLeuG 422
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
657 GCTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 706
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
422 nLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 439
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
707 AG.....GCTTGGAGGCCAGCCACCAATTCAGCAGCAGCAG 744
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
439 InProGlnAlaThrThrLeuThrAlaProGlnProGlnValProPro 455
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
745 ATGACAGCAGCAGCAGCAGCTCCGCTCCAGGCTGCGCCAGCAGCAG 794
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
456 ThrGln.....GlnValProProSerGlnSer.....G 465
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
795 GCAGATGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 844
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
465 nGlnGlnAlaGlnThrLeuValValGln..... 474
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
845 CTCAGATTGCTCAGAACCAACCATCACAACCTCCGCCAGCTGCCAGAC 894
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
475 ..PrometLeuGlnSerSerProLeuSerLeuProProAlaLeuAlaPro 490
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
895 CAGCCTTGGTGTCCAGAGGCCAGAGCTCCTCCGACAAATGTTGTATAC 944
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
491 LysProProIlePro.....IleGlnSe 498
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
945 CCAACCCAGCAGTGAATTTGTCGAGCTCGAGTGGTG..... 981
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
498 rLysProProValAlaProIleLysProProGlnLeuGlyAlaAlaLysM 515
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
982 .....GTGCAGCAGCCCGCCAGTGGCAGCCCGCCAGTGCAGCAGCAG 1023
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
515 etSerAlaIleGlnIleProPro.....ProHisIleProValGlnVal 529
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1024 ACAGCAGTACAGCAGCTCAGGCTGCCAGATGAGTGTCCCGGAGTCCA 1073
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
530 ValGlyThrArgGlnProGlyThrAlaGlnAlaGlnAlaLeuGlyLeuAl 546
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1074 GGTCAAGCAGCAGCAGCTCCCGCTGTCTCCGCTGACCGGGCCAGC 1123
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
546 aglnLeuAlaAlaValPro..... 553
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1124 AGGTGACAGCCCGCAGTGCATGCCCTCCCGCCAGCCGCGCCAG 1173
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
554 .....ThSerArgGlyMetProGlyThrValGln...SerGlyGln 566
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1174 CCGGCGCAGCCAGCTCAGCAGCCCAACTCACAAGTCAAGTCAAGTCCG 1223
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
567 AlaHisLeuAlaSerSerProProSerSerGlnAlaProGlyAlaLeuG 583
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1224 CCGATCTCCAGTACTGCTCTCTCCAGCCCTCAGCCGAGCCCTCCAG 1272
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
583 nGlySerProProThrLeuAlaProGlyMetThrLeuAlaProValGlnG 600
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1273 .....AGCCAGTGAAG 1284
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
600 lYThrAlaHisValValLysGlyGlyAlaThrThrSerSerProValVal 616
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1285 GCGCGGACCCCAAGAACTTCAGTCCCTGACCTGAGCTTTAAACAC 1334
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
617 AlaGlnValProAlaAlaPheGlyMetGlnSer.....ValHisLe 630
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1335 ACCTGTGAACCCAGCAGCTGTCTGATGAGCCAGCTGGCTCCAGCAGCTG 1384
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
630 uProGlyLysProGlnThrLeu.....AlaValLysArgLysAlaA 644
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1385 AGGAGCAGCAGTACTGGACAGCTGAAGCAGCTGCTCGAAGTCAATCGAG 1434
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
644 sPserGlnGlnGlnIuArgAspAspValSerThrLeuGlySerMetLeuPro 660
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1435 CCGCTCGCCGATGATCAACAAGATTCGACAGAAGCAAGCAAGAAAAA 1484
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
661 AlaLysAlaSerProValAlaGlnSerProLysValMetAspGlnLysSe 677
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1485 GGACCTG...AGTAGATGAAGACCTTGTGACAT..... 1518
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
677 rSerLeuGlyGlyLysAlaGlnSerValAlaAsnValAsnAlaAsnThrP 694
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1519 .....CTGACAGACCCCTCGAAGCGGTGCTCCCTG 1548
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
694 roSerSerGlnLeuValAlaLeuThr..ProAlaProSerValProProP 710
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1549 AAGACCTTGCAAAAGTGTGAGATCGCCCTGAGAAACTCAGAAATGACAT 1598
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
710 roThrLeuAlaMetValSerArgGlnMetGlyAspSerLysProProGln 726
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1599 GCGGTGCCACCTGCC 1613
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
727 AlaIleValLysPro 731
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /STD2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW52830
seq_documentation_block:
ID AAW52830 standard; Protein; 1004 AA.
XX
AC AAW52830;
XX
DT 09-JUL-1998 (first entry)
XX
DE A tumour suppressor gene called polyhomeotic 1 (hph1).
XX
KW Tumour suppressor gene; polyhomeotic 1; hph1; human chromosome 12p13;
XX identification; neoplastic tissue; cellular differentiation; diagnosis;
XX neoplasia; ss.
XX
OS Homo sapiens.
XX
PN W09807858-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-US14866.
XX
PR 06-FEB-1997; 97US-0036939.
XX 23-AUG-1996; 96US-0024349.
XX 04-DEC-1996; 96US-0031569.
XX
PA (CHIR ) CHIRON CORP.
XX
PT Randazzo F;
XX
DR WPI; 1998-169162/15.
XX
DR N-PSDB; AAV21060.
XX
PT Isolated human polyhomeotic 1 oncogene - used to develop products
XX for diagnosis and therapy of proliferative and developmental
XX disorders, e.g. neoplasia, dysplasia or hyperplasia
XX
PS Claim 1; Pages 28-30; 40pp; English.
XX
CC The present sequence encodes a novel human tumour suppressor gene
XX termed polyhomeotic 1 (hph1). The hph1 gene maps to human chromosome
XX CC

```

12p13, a region which is frequently lost in non-small cell lung cancer and breast cancer. A method of identifying neoplastic tissue of a human comprises comparing the expression of a hpl1 gene in a tissue of a human suspected of being neoplastic with the expression of a hpl1 gene in a tissue of the human which is normal. Under-expression of the hpl1 gene identifies the subject as having neoplastic tissue. The hpl1 oncogene functions to suppress neoplasia and dysplastic or hyperplastic cell growth as well as to induce cellular differentiation. The cDNA, protein and vectors can be used as diagnostic and therapeutic tools for cell proliferative and developmental disorders and to identify a p13 region of a human chromosome 12. They can be used for the detection, diagnosis or prognosis of neoplasia or for detecting a genetic predisposition to neoplasia. They can also be used to treat tumours.

Sequence 1004 AA:

alignment_scores:
Quality: 389.00 Length: 522
Ratio: 1.379 Gaps: 20
Percent Similarity: 54.023 Percent Identity: 28.544

alignment block:
us-09-668-119-1 x AAW52830 ..

Align seg 1/1 to: AAW52830 from: 1 to: 1004

```

178 CAGAGCCCTGGAGTGGAGCTGGCGGAGCCCGCTGGAATGGCATGCC 227
    |||||||.....:|||||:|||||:|||||:|||||:|||||
260 GlnserLeuasnleuSerGlnAlaGlyGlySerGlyAsnSerIlePr 276
    |||||||.....:CTCGGGGAGCCCGGAGAGTCTGTGGGGATGGGTA 285
228 T.....:|||||:|||||:|||||:|||||:|||||:|||||
276 OGlySerMetGlyProGlyGlyGlyGlnAlaHisGlyGlyLeuGly 293
    |||||||.....:GCTTGGTGGCATGGGACAGCCATGCTCTGAGGGACGGCCCTCT 315
266 GCTTGGTGGCATGGGACAGCCATGCTCTGAGGGACGGCCCTCT 315
    :|||:|||||:|||||:|||||:|||||:|||||
293 LIneuProSerSerGly.....MetGlyGlySerGlyProAlaGly 307
    :|||:|||||:|||||:|||||:|||||:|||||
316 GGGAGCTCGGGGATGGCCCTCACAGAGCTGCTGCTGCTGAGGAGC 365
    |||||||:|||||:|||||:|||||:|||||:|||||
308 GLyThrGlyAlaValAlaGlnProLeuProAlaAlaGlnThrVal 324
    :|||:|||||:|||||:|||||:|||||:|||||
366 TCACAGAGCCAGCTGAGCTCCAGAGAGTGGCGCTGCAGAGCAGC 415
    :|||:|||||:|||||:|||||:|||||:|||||
324 rGlnGlySerGlnThrGlnAlaGlnSerAlaAlaAlaGlyAlaGln 341
    :|||:|||||:|||||:|||||:|||||:|||||
416 AACAGGAGCAGTTCACAGAGCAGCAGCGCGCTACAGAGCAGCAG 465
    :|||:|||||:|||||:|||||:|||||:|||||
341 LAsnProGlySerGlyGlnGlnAlaHisValGlyMetAsnLeuThr 357
    :|||:|||||:|||||:|||||:|||||:|||||
466 CAGAGAGCAGCAGCAGCAGTTCACAGAGCTCACAGAGTGCATGCAG 515
    :|||:|||||:|||||:|||||:|||||:|||||
358 ThrProAlaProSerGlnThrLeuIleSerSerAlaThrThyThrGln 374
    :|||:|||||:|||||:|||||:|||||:|||||
516 GCGATTCACAGAGCAGTTCAGAGCAGCAGAGCTCCAGAGCAGCAG 565
    :|||:|||||:|||||:|||||:|||||:|||||
374 eGlnProHisSerLeuIleGlnGlnGlnGlnGlnGlnGlnGlnGln 391
    :|||:|||||:|||||:|||||:|||||:|||||
566 AGCAGCAG.....CAGCATCTAATTAATTCATCAATAATGAGCA 609
    :|||:|||||:|||||:|||||:|||||:|||||
391 yAsnValValIleGlnGlnGlnGlnIleAlaIleHis.....GlnGln 405
    :|||:|||||:|||||:|||||:|||||:|||||
610 CAGATTCAGAGCAGCAGCAGCAGTTCAGAGCAGCAGTTCAGAGCAG 656
    :|||:|||||:|||||:|||||:|||||:|||||
406 GlnProGlnHisArgGlnSerGlnLeuHisThrAlaThrHisLeuGln 422
    :|||:|||||:|||||:|||||:|||||:|||||
657 GCTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 706
    :|||:|||||:|||||:|||||:|||||:|||||
422 nLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 439
    :|||:|||||:|||||:|||||:|||||:|||||
707 AG.....GCTTGGAGGCCAGCCCAATTCAGAGCAGCAGCAG 744
    :|||:|||||:|||||:|||||:|||||:|||||

```

```

439 InProGlnAlaThrThrThrLeuThrAlaProGlnProProGlnValProPro 455
    |||
745 ATGCAGAGCAGCAGAGCTCCGCCCTCCAGAGCTGCCCCAGAGCTGCA 794
    |||
456 ThrGln.....GlnValProProSerGlnSer.....Gln 465
    |||
795 GCGATGATGATACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 844
    |||
465 nGlnGlnAlaGlnThrLeuValValGln..... 474
    |||
845 CTCAGTCTGCTGAGACCAACCATGCACAGCTCCGCCAGTCGAGCAGCAG 894
    |||
475 ..ProMetLeuGlnSerSerProLeuSerLeuProProAlaAlaPro 490
    |||
895 CAGCCTTTGGTGTCCACAGCGCAGAGCTCTCCCTGAGCAAAATGTTGTAAC 944
    |||
491 LysProProIlePro..... 981
    |||
945 CCAACCCAGCAGTGAATTTGTCGAGCTCCGATGGTG..... 981
    |||
498 rLysProProValAlaProIleLysProProGlnLeuGlyAlaAlaLysM 515
    |||
982 .....GTCCAGCAGCCCCCAGTCGAGCAGCAGCAGCAGCAGCAGCAG 1023
    |||
515 eSerAlaAlaGlnGlnProPro.....ProHisIleProValGlnVal 529
    |||
1024 ACAGCAGTACAGACAGCTCAGGCTCCGAGATGCTGCTCCCGAGTCCA 1073
    |||
530 ValGlyThrArgGlnProGlyThrAlaGlnAlaGlnAlaLeuGlyLeuAl 546
    |||
1074 GGTACAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1123
    |||
546 aGlnLeuAlaAlaAlaValPro..... 553
    |||
1124 AGGTGACAGCAGCAGCAGTGCATGCTCCGCCCTCCAGAGCTCCGCCAG 1173
    |||
554 .....ThrSerArgGlyMetProGlyThrValGln..SerGlyGln 566
    |||
1174 CCGGCGCAGCAGCAGCTCAGAGCCCACTGCAAGCTGAGCTGAGCCTGCG 1223
    |||
567 AlaHisLeuAlaSerSerProProSerSerGlnAlaProGlyAlaLeuGln 583
    |||
1224 CCCATCTCCCACTGAGTCTGCTCCGCCAGCAGCAGCAGCAGCAGCAG 1272
    |||
588 nGlnLysProProThrThrLeuAlaProGlyMetThrLeuAlaProValGln 600
    |||
1273 .....AGCCAGTGCAG 1284
    |||
600 LyThrAlaHisValValValLysGlyGlyAlaThrThrHisSerProValVal 616
    |||
1285 GCGGAGCAGCAGCAGCAAGTTCAGTGTCCCTCAGCTGAGCTTTAAACAC 1334
    |||
617 AlaGlnValProAlaAlaProMetGlyMetGlnSer.....ValHisLe 630
    |||
1335 ACCTGTGAACCCAGCTGTGTCATGAGCCAGCTGAGCTGAGCTGAGCAGCAG 1384
    |||
630 uProGlyLysProGlnThrLeu.....AlaValLysArgGlyAla 644
    |||
1385 AGGAGCAGCAGTACCTGAGCAGCAGTGAAGCAGCTGTGAAATGATCAGC 1434
    |||
644 spSerGlnGlnGlnArgAspArgValSerThrLeuGlySerMetLeuPro 660
    |||
1435 CCGCTGCGCGCATGATCAACAGATGCAGACAGACAGACAGACAGACAG 1484
    |||
661 AlaLysAlaSerProValAlaGlnSerProLysValMetAspGlnLysSe 677
    |||
1485 GGACCTG...AGTAAGATGAAGACCTTCTGAGCAT..... 1518
    |||
677 rSerLeuGlyLysLysAlaGlnSerValAlaAsnValAlaAsnAlaThr 694
    |||
1519 .....CTGACAGACCCCTCGAAGCGGTGCTCCCTG 1548
    |||
694 rSerSerGlnLeuValAlaLeuThr..ProAlaProSerValProPro 710
    |||

```



```

CC and Rb proteins. The cellular proteins of the invention are used for
CC in vitro screening of compounds (including immunosuppressants) which
CC are able to interact physically and/or functionally with them,
CC particularly to determine sensitivity of the protein to
CC immunosuppressants or to identify capacity to interact with
CC Ela viral proteins. The proteins can be used to induce (or interfere
CC with) cellular differentiation and proliferation, including formation
CC of malignant cells.
XX
XX
SQ      Sequence      3190 AA;

alignment_scores:
      Quality:      342.00      Length:      600
      Ratio:      1.089      Gaps:      25
Percent Similarity: 52.333      Percent Identity: 26.167

alignment_block:
US-09-668-119-1 x AAB84634 ..

Align seg 1/1 to: AAB84634 from: 1 to: 3190

34  AGCAGGATATGGAGAGCCATGTTTCTGGAAGCCAGCCGCGAGA 83
      ::::::::::: 342.00      ::::::::::: |||::|::|::|::|
      Ratio: 1.089      Gaps: 25
984  AAGAGLTVaIGLTVaSGPMETTYrGLMETAlaLVsSERATySGEGL 1000
      ::::::::::: 52.333      ::::::::::: :::::::::::
84  ATACCTTCTCTCGTGGCCAGGCTCATTTACATTTT...CGAGCATTC 130
      |||  ::|::|::|::|::|  ||  ::  :::::::::::
1000  uTYrTYrHsLeuLaIGLTVaSLTYrYSLIeGLnYSGLueG 1017
      ::|::|::|::|::|::|::|  ::  :::::::::::
131  ATACACAGAAA.....TCT 144
      ::|::|::|::|::|::|::|  ::  :::::::::::
1017  InGLnLYsArGLueuLYsArGLySLnGLnHsGLnGLMeLeuMeGLn 1033
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
145  CAAGCTCCGTCATGCATCATATGATGACTCCAGAGCCTCAGCGCG 194
      |||  ::|::|::|::|::|::|::|::|::|::|::|::|::|
1034  GLnGLnGLyValAlaAsnProValaGLyGLyAlaLaGLyGLyAlaGL 1050
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
195  ACCCTGTGCGGGA...GCCGCTGAATTGGCATGCTCTCGGGGCCGG 241
      |  |||::|::|::|::|::|::|::|::|::|::|::|::|::|
1050  ySERAlaLaGLyValaLaGLyGLyValaLeuProGLnGLnGLnGLnG 1067
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
242  GACAG..... 246
      |||
1067  InGLnGLnGLnGLnGLnGLnGLnGLnGLnGLnProLeuGLnSerys 1083
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
247  .....TCTGCGCGGATGCTAGCTTGTGTCGCCATGGAGCACC 287
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
1084  lLHsPProSerTlLeSerProMeGLy....GLYAlMet...ProPr 1097
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
288  AATGTCTCTCTCAGGGCAGCCGCTCCTCGGACCTGGGGAGT..... 330
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
1097  oGLnGLnLeuArGProGLnGLyProProGLyLlLeuGLnGLnGLnThA 1114
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
331  .....GCCCTCAGACAGATGGCTGCGG 354
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
1114  lAaLaGLyLeuGLyValGLyValGLyAlaThrAsnAsnMeValAlThrMeT 1130
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
355  TCTACGCAACTCCACAGACCAGCTGAGCTCCAGCAGGTGGCGCTGA 404
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
1131  ArGserTlLeSerProGLyGLyAsnMeT.....LeuAlaLeuGL 1143
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
405  GCACAGCAGCAACAGCAGCAGCTTCCAGCAGCAGCAGCAG..... 444
      |||::|::|::|::|::|::|::|::|::|::|::|::|
1143  nGLnGLnGLnArGMeTGLnPhEProGLnGLnGLnGLnGLnProGL 1160
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
444  ..... 444
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
1160  LySerGLyAlaGLyLySMeTLeuValGLyProProGLyProSerProGLy 1176
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
445  .....GCGGCGCTACAGCAGCAGCAGCAG..... 468
      |||::|::|::|::|::|::|::|::|::|::|::|::|
1177  GLyMeTValValaAsnProAlaLeuSerProTYrGLnThrThrAsnValle 1193

```

```

469 .....CAGCAGCAACAGCAGCTTC..... 489
1193 uThrSerProValProGlyGlnGlnGlnGlnPheIleAsnAla 1210
490 .....CAGGCTCAGCAGAACTGCCATGCAGCAGCAGTTCC 522
1210 snGlyGlyThrGlyAlaAsnProGlnIleuSerGluIleMetLysGln... 1225
523 CAGCAGTACTGTCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCA 572
1226 ArgHisIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1242
573 GCAGCATCTAATT.....AAATTGCATC 595
1242 nGlyMetLeuPProGlnSerProPheSerAsnSerThrProLeuGln 1259
596 ATCAAAATCAGCAGACAGTACAGCAGCAGCAGCAGCAGCAGCA 645
1259 lnglnglnglnglnglnglnglnglnglnglnglnglnglnglng 1275
646 GCACAGCTGCAGCTCCAAACAGCAGCAGCAGCAGCAGCAGCA 695
1276 SerPheSerSerProMetGlnGlnGlnGlnGlnGlnGlnGln 1292
696 GCAGCAGCA..... 704
1292 nGlnLysProGlySerValLeuAsnAsnMetProProThrProThrSerL 1309
705 .....GCAGGCTTTGGAGCCGAGCCGAGCCGAGCCGAGCCG 740
1309 euGlnAlaIleuAsnAlaGlyAlaGlyAlaProGlyThrGlyLysAla 1325
741 ACCGATGCAGCAGCCAGCAGCC...TCCGCCCTCCAGGC.....TC 778
1326 SerAsnValThrValSerAlaProSerProSerProGlyPheLeuSerAs 1342
779 TGCCCGCAGCAGTCGACGAGATGATCAGCAGCAGCAGCAGCCG 828
1342 nGlyProSerIleGlyThrProSerAsnAsnAsnAsnSerSerSerAla 1359
829 CCACAGCCCA.....GCAGCCTCCAGTTGCTCAGACAAACATCA 872
1359 snAsnAsnProProSerValSerSerLeuMetGlnGlnProLeuSerAsn 1375
873 ACTCCGCCCA.....GTCCAGAGCC 895
1376 ArgProGlyThrProProGlyIleProAlaSerProValProAlaThrSe 1392
896 AGCCTTGGTGTGCAGCAGCAGCAGCTCCCTGGACAATGTTGTAATCC 945
1392 rAlaSerGlyLeuAlaAlaSerSerThrPro.....A 1403
946 CAACCCACTGAATTTGTCAGAGCTCGATGGTGAGCAGCAGCCCC 995
1403 laSerAlaAlaAlaThrCysAlaSerSerGlySerIleSerAsnSerSer 1419
996 AGTGACCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCAG 1045
1420 SerGlyAlaThrAlaAlaGlyAlaSerSerThrSerSerSerSer.. 1435
1046 CTGCCAGATGTGTGCTCCGAGTCA...GGTCAGCAGCAGCAGCTC 1092
1436 .....AlaGlySerGlyThrProLeuSerSerValSerThrProT 1449
1093 CCATATCTGCTCCGCGCTCAGCCGGGCAAGCAGTGATGACCCGCACTC 1142
1449 hSerIleAlaThrMetAlaThrSerSerGlyGlyGlyGlyGlyGly 1465
1143 GATGCC...CCCTCCCGCCAGCCGTCGCCGAGAGCCGCGCCAGCCAGCT 1189
1466 AsnAlaGlyGlyGlySerSerThrThrProAlaSerAsnProLeuLeu 1482

```

```

1190 CACAGCCCACTCCACACTCACTCTGCG..... 1218
1482 uMetSer GlyGlyThrAlaGlyAlaGlyThrGlyAlaThrThrThr 1498
1219 CCTGCCCATCTCCGAGTACTCTCCGAGCCGAGCCGAGCCGCTC 1268
1499 SerThrSerSerSerSerArgMetMetSerSerSerSerSerLeu 1515
1269 CCAGAGCCCAATGACGGCGGCGGAGCCAGCAACTTCACTGCTCCCTC 1318
1515 rGlnMetAlaAlaLeuGlnAlaAlaAlaAlaArgAspAsnAspGluThr 1532
1319 CTGACCTTTAAACACACCTGTGAAACCCAGCTCTGTGATGACCCAGCT 1368
1532 roSerProSerGlyGlnAsnThrAsnGlySerGlyGlySerGlyAla 1548
1369 GGCCTCAGCCAGGCTGAGGAGCAGTACCTGATGACAGCAGCAAG 1416
1549 GlyGlyMetAlaSerLysGlyLys.....LeuAspSerIleLysGln 1562

seq_name: /STD2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV74791
seq_documentation_block:
ID AAV74791 standard; Protein; 233 AA.
XX
XX AAV74791;
XX
XX DT 21-MAR-2000 (first entry)
XX
XX DE Neisseria meningitidis ORF 263 protein sequence SEQ ID NO:1056.
XX
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX OS Neisseria meningitidis.
XX
XX PN W09957280-A2.
XX
XX PD 11-NOV-1999.
XX
XX PF 30-APR-1999; 99W0-US09346.
XX
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Piza M, Rappunoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tetteijn H, Venter JC;
XX
XX DR WPI: 2000-062150/05.
XX DR N-PSDB; AA253553.
XX
XX PT Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
XX PS Claim 2; Page 606; 1453pp; English.
XX
XX CC AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
XX CC PCR primers used in the exemplification of the present invention. The
XX CC polypeptides, the polynucleotides, antibodies and compositions of
XX CC the invention can be used as vaccines, as diagnostic reagents, and as

```


immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to Neisserial bacteria (e.g. meningitis and septicemia), to detect the presence of Neisseria bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

Sequence 233 AA;

alignment_scores:

Quality: 341.00 Length: 257
Ratio: 2.706 Gaps: 15
Percent Similarity: 49.027 Percent Identity: 31.128

alignment_block:

US-09-668-119-1/rev x AAY74791 ..

Align seg 1/1 to: AAY74791 from: 1 to: 233

```

1004 GCGTCGACATGGGGCTGC.....TGACACACATCGAGCTCGGACAAA 961
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 glyCysAlaGlyGlyCysGlyAlaAlaAlaThrThrThrGlyAla...CysG1 19
960 TTTCAGTGGTGGTGGTATACACATTTGTCACGAGGAGAGCTTGGCT 911
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 yAlaThrGly.....CysCysAlaAlaAlaCysThrC 30
910 GTGACACCAAGGCTGGGTGCTGCGACTGTGGGGGAGTGTGATGATGG 861
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 ySg1yGly.....CysGlyCysGlyCysThrCys..... 39
860 TTCTGACGACACTGGAGCTGCTGCTGCTGGGCTGGCTGGTGGTGGT 811
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 .....GlyCys.....CysGly.....CysCys 45
810 TGTGTGATCATCTGCTGC.....AGCTGCTGGGAGAGAGCTGGTGG 770
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 sThrThrCysAlaCysCysCysAlaAlaGlyCysGlyThrAlaAla 62
769 AGGGC.....GGAGCTGTGGCTGC... 750
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 hGlyGlyCysGlyAlaAlaAlaAlaAlaGlyCysGlyCysGly 78
749 .....TGCATGGTGGCTGCTGAATTGGTGGCTGGGCTGCCAA 712
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 GlyThrAlaThrCysCysGlyAlaCysGlyAlaGlyGly.....Al 92
711 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 aAlaCysThrCysAlaAlaAlaAlaGlyCysAlaThrThrThrThrCysG 109
661 GGAGCTGACGCTGTCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 lAlaAlaThrGlyCysGlyGlyCysThrAlaCysAlaAlaCysGysAla 125
611 TGTGTGATTTTGTGATGATGCAATTTAATAGATCTGCTGCTGCTGCT 562
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126 GlyCys.....AlaGlyCysAlaGlyCysAlaGlyThrCys 138
561 C..... 561
138 sGlyAlaAlaAlaGlyThrCysGlyThrGlyAlaAlaThrGlyGlyCysGlyT 155
560 .....TGCTGCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 522
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 hAlaSerThrCysThrGlyGlyCysAlaAlaCysCysCysThrGlyThr 171
521 AACTGCTGCTGCATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 472
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 GlyCysAlaAlaCysThrAlaCys.....GlyThrCysAlaAlaCysAl 186

```

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB95124

seq_documentation_block:
ID AAB95124 standard; Protein: 307 AA.

```

XX AAB95124;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:17122.
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17122; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX in gene therapy. The primer sets can be used in antisense therapy and
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

```

CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SO Sequence 307 AA:

alignment scores:

Quality: 336.50 Length: 317
 Ratio: 2.588 Gaps: 19
 Percent Similarity: 41.009 Percent Identity: 27.445

alignment block:

US-09-668-119-1/rev x AAB95124 ..

Align seg 1/1 to: AAB95124 from: 1 to: 307

```

1133 GTCTGCACCTGCTGCGCCGGTGACGGCGAGACACATGCGCGAGCTGCT 1084
      ::::::::::::::::::::
11   IleCysValCysValPro.....ValCysValGlyAlaHisLe 23
1083 CTGGCTGACCTGGACTCCGGAGCGACACCATGCTGCGCGAGCTGCT 1034
      |
23   U.....CysValTrrPvalHisIleCysValC 32
1033 GTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1005
      ::::::::::::::::::::
32   ysValAlaGlyLeuCysGlyCysThrPheValCysValCysAlaCysValCys 48
1004 GGCCTGCACCTGGGGCTGCTGACACCATGCGAGCTGGCAATTTTCAG 955
      ::::::::::::::::::::
49   AlaCysValCysGlyCysLeuCysVal..... 57
954 TGGGTGGGTATACAACTTTGTCACGAGAGAGCTTGGCGCTGTCACA 905
      |||
58   .....TrrPglyThrPheValCys.....ValCysAlaCys.... 67
904 CCAAGAGCTGGGTC.....TGCAGCTGCGCGAGAGCTGCTGCTGCTGCTG 861
      |||||
68   ....ValTrrPvalHisIleCysValCysAlaSerValCys..... 79
860 TTCGACGACACTGAGGCTGCTGCGGCTGCTGCGGCTGCTGCTGCTGCTG 811
      |
80   .....ValProValCysGlyAlaHisLeuCysVal 90
810 TGTGTGATGC.....ATCTGC...TGCAGCTGCTGCGGCGAGAG 776
      |||
90   lArGleuCysValGlyAlaHisIleCysValCysAlaCys..... 103
775 CCTGGAGAGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 747
      |||||
104   .....ValCysGlyCysThrPheValCysValCysAlaCys 115
746 ATCGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 700
      ::::::::::::::::::::
116 ValCysGlyCysLeuCysValCysGlyAlaHisLeuCysValCysValCy 132
699 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 665
      |
132 sleuCysValGlyAlaHisLeuCysValCysLeuCysValCysAlaCysV 149
664 GTTGGAGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630
      |||||
149 alTrrPglyCysThrPheValCysValCysAlaCysValCysAlaCysValCys 165
629 .....TGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 569
      |||
166 PheValCysValCysAlaCysValCysGlyCysThrPhe..... 178
588 TTTATTAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
      ::::::::::::::::::::
179 .ValCysValCysAlaCysValCysGlyCysThrPheValCysValCysL 195
538 GCTGCACTACTGCTTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 489
  
```

```

1195 eucCysValTrrPvalHisIleCysValCysValCysLeuCysValTrrPval 211
      ::::::::::::::::::::
488 AACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 454
      ::::::::::::::::::::
212 HisIleCysValCysAlaCysValTrrPvalHisIleCysValCysValPr 228
453 TAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404
      ::::::::::::::::::::
228 oValCysValCysLeuCysValTrrPvalHisIleCysValCysValCysL 245
403 GCAGCGCACCTGCTGAGACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTG 354
      ::::::::::::::::::::
245 eucCysValTrrPvalHisIleCysValCysValCysAlaCysValCys... 260
353 ACGACAGCATGCTGAGGGGCGCATCCCGAGAGTCCAGAGGCGGCTG 304
261 .....ValAlaGlyProGlyCys 266
303 C 303
266 s 266
seq.name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV81609
seq_documentation_block:
ID   AAV81609 standard; Protein: 1237 AA.
XX   AAV81609:
XX   24-MAY-2000 (first entry)
XX   Streptococcus pneumoniae type 4 protein sequence #109.
XX   DE
XX   Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX   KW   Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX   KW   antibacterial; antiinflammatory; meningitis; infection; diagnosis;
XX   KW   pneumococcal disease.
XX   OS   Streptococcus pneumoniae.
XX   PN   WO200006737-A2.
XX   PD   10-FEB-2000.
XX   PF   27-JUL-1999; 99WO-GB02451.
XX   PR   27-JUL-1998; 98GB-0016337.
XX   PR   19-MAR-1999; 99US-0125164.
XX   PA   (MICR-) MICROBIAL TECHNIQUES LTD.
XX   PI   Glibert CFG, Hansbro PM;
XX   DR   WPI: 2000-195300/17.
XX   PT   New Streptococcal protein, useful as a vaccine, for diagnosis of
XX   PT   pneumococcal diseases and for screening agents capable of antagonizing
XX   PT   or inhibiting expression of the protein
XX   PS   Claim 1; Page 83; 108pp; English.
XX   AA
XX   AAV81501 to AAV81679 represent specifically claimed protein sequences
XX   isolated from Streptococcus pneumoniae. AAV81407 to AAV81590 represent
XX   specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX   The sequences have antibacterial and antiinflammatory properties.
XX   The protein sequences, and fragments of them, are useful as immunogens
XX   and/or antigens. The nucleotide sequences can be used in vaccines and in
XX   diagnostic assays. The proteins and nucleotides can be useful for the
XX   detection and diagnosis of S. pneumoniae. The protein sequences are also
XX   useful for screening an agent capable of antagonising, inhibiting or
XX   interfering with the function or expression of the proteins in which the
XX   agent is useful for treatment or prophylaxis of S. pneumoniae infection
XX   and meningitis. AAV81591 to AAV81614 represent primers used in the
  
```


ID AAB69609 standard; Protein; 121 AA.
AC AAB69609;
XX
XX
DT 30-APR-2001 (first entry)
XX
DE Huntington accumulation inhibitor peptide HD-Q104-GFP.
XX
XX Neurological disorder; Huntington's disease; Alzheimer's disease;
KW Parkinson's disease; prion disease; frontotemporal dementia;
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW dentatorubral-pallidoluysian atrophy; spinocerebellar ataxia type 1;
KW SCA2; SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
OS Synthetic.
XX
PN WO200106989-A2.
XX
PD 01-FEB-2001.
XX
PF 24-JUL-2000; 2000WO-US20131.
XX
PR 27-JUL-1999; 99US-0146047.
XX 21-JUL-2000; 2000US-0620955.
XX
XX (HUST/) HUSTON J S.
PA (MESS/) MESSER A.
PA (LECE/) LECERF J.
XX
PI Huston JS, Messer A, Lecerf J;
XX WPI; 2001-182700/18.
DR
XX
XX Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody -
XX
XX Disclosure; Page 98; 108pp; English.
XX
XX The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubral-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7.
CC
XX
SQ Sequence 121 AA:

alignment_scores: Quality: 328.50 Length: 111
 Ratio: 4.056 Gaps: 1
Percent Similarity: 72.973 Percent Identity: 69.369

alignment_block:

US-09-668-119-1 x AAB69609 ..

Align seg 1/1 to: AAB69609 from: 1 to: 121

```
376 CAGCTGACGCTCCAGACGAGTGGCGCTGACAGCAGCAGCAGCAACAGCAGCA 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18 GInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 34
426 GTTCAGCAGCAGCAGCAGCGCGGCTACAGCAGCAGCAGCAGCAGCAGCAGC 475
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 nGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 51
476 AACAGCAGAGTTCAGAGCTCAGCAGAGTGCATGACGAGCAGCAGCAGTTC 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 InGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 67
```

```
526 GCAGTAGTCAGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCA 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 GInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 84
576 GCATCTAATTAATTGCATCATCAAAATCAGCAGCAGATACAGCAGCAGC 625
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 n.....GInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 94
626 AACAGCAGCTGACAGCCGAAATAGCAGAGCTGACGCTCCAAACAGCAGCAG 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 InGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 110
676 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 GInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 121
```



```

368 ..GGAGTTGCCGTAGACACGACGACCATGCTGTAGGGGGCCATCCCCAG 321
      |||::|||      :::      |||:::
627 hrglyAlaIaIaCysThrGlyCysGlyGlyCysThrGlyCysThrAlaThr 643
      :: |||::|||:::      |||::|
320 GTCCAGGAGGCGGCTGCCCTGAGAGAGACATTTGGCTTCCCATGGCAC 271
      :: |||::|||:::      |||::|
644 ThrThrGlyGlyGlyAla.....GlyCysAla..... 652
      |||::|||:::      |||::|
270 AAAGTACCATCCCGCCAGAGACTGTCCCGGGCCCGAGAGAGC 225
      |||::|||:::      |||::|
653 .....CysGlyGlyAlaIaCysGlyGly 659
      |||::|||:::      |||::|

seq_name: /cgn2_6/prodata/2/iaa/5A_COMB.pep:US-07-906-349A-6

seq_documentation_block:
; Sequence 6, Application US/07906349A
; Patent No. 5434064
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; APPLICANT: Skolnik, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARIOTIC TYROSINE KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906,349A
; FILING DATE: 30-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/643,237
; FILING DATE: 18-JAN-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-906-349A-6

alignment_scores:
      Quality: 446.50      Length: 594
      Ratio: 1.892      Gaps: 31
Percent Similarity: 39.731      Percent Identity: 25.253

alignment_block:
US-09-666-119-1/rev x US-07-906-349A-6 ..

Align seq 1/1 to: US-07-906-349A-6 from: 1 to: 801

1600 GCATGTCATTTGTGAGTTTCACAGGCGGATGTCACACTTTGGAGGTC 1551
      |||::|||:::      |||::|
197 AlaGlyShs.....GlyGlnThrGlyMetPheProArgAs 208
      ::|||::|||:::      |||::|
1550 TTCAGGGGACCCGCTTCGAGGGCTGTCTCAGAAATGTCCAGAAAGGCTTT 1501
      ::|||::|||:::      |||::|

```

[illegible]

```

1 TITLE OF INVENTION: TUMOR SUPPRESSOR
2
3 NUMBER OF SEQUENCES: 2
4
5 CORRESPONDENCE ADDRESS:
6
7 ADDRESSEE: CHIRON CORPORATION
8
9 STREET: 4560 HORTON STREET
10
11 CITY: EMERYVILLE
12
13 STATE: CA
14
15 COUNTRY: USA
16
17 ZIP: 94608
18
19 COMPUTER READABLE FORM:
20
21 MEDIUM TYPE: Floppy disk
22
23 COMPUTER: IBM PC compatible
24
25 OPERATING SYSTEM: PC-DOS/MS-DOS
26
27 SOFTWARE: PatentIn Release #1.0, Version #1.30
28
29 CURRENT APPLICATION DATA:
30
31 APPLICATION NUMBER: US/08/916,352
32
33 FILING DATE:
34
35 CLASSIFICATION: 435
36
37 ATTORNEY/AGENT INFORMATION:
38
39 NAME: POTTER, JANE
40
41 REGISTRATION NUMBER: 33,332
42
43 REFERENCE/DOCKET NUMBER: 1355.
44
45 TELECOMMUNICATION INFORMATION:
46
47 TELEPHONE: 510-923-2707
48
49 TELEFAX: 510-655-3342
50
51 INFORMATION FOR SEQ ID NO: 2:
52
53 SEQUENCE CHARACTERISTICS:
54
55 LENGTH: 1004 amino acids
56
57 TYPE: amino acid
58
59 STRANDEDNESS: single
60
61 TOPOLOGY: linear
62
63 MOLECULE TYPE: protein
64
65 US-08-916-352-2

```

Alignment_scores:	Quality:	389.00	Length:	522
Ratio:	1.379	Gaps:	20	
Percent Similarity:	54.023	Percent Identity:	28.544	
alignment_block:				
US-09-668-119-1 x US-08-916-352-2 ..				
Align seg 1/1 to: US-08-916-352-2 from: 1 to: 1004				
178	CAGACGCTGATGGGCGGACGCTGCTGCGGAGCGCGGTGGATTTGGCATGGC	222		
266		266		
260	GlnSerLeuAnIeuSerGlnAlaGlyGlyGlySerGlyAnSerIlePr	276		
228	T...P.....CTCGGGGCGCGGACAGTCTCTGAGCGGATGGGTA	265		
276	oGlySerMetCylProGlyGlyGlyGlyGlnAlaGlnIleGlyGlyLeuGlyG	293		
266	GCTTGGTGGCATGGAGACGSCATTTGCTCTGTAGGGACGCGCGCTCT	315		
293	IlnleuProSerSerGly.....MetGlyGlyGlySerCysProGlyGlyS	307		
316	GCGACCTCGGGGATGGGCGGCTCACAGCATGAGTCTGCTGTACGCGCAC	365		
308	GlyThrGlyValValGlnProLeuProAlaAlaGlnThrValThrValSe	324		
366	TCCACAGACGCCAGTGGAGCTCCAGCGAGGTGGGCGCTGCAGCAGCAGC	415		
324	TGlnGlySerIlnThrGlyIlnIleGlnSerAlaAlaAlaGlyAlaGlnIla	341		
416	AACAGCAGCAGATTCTACAGCAGCAGCAGCGCGGCGTACAGCAGCAGCAG	465		
341	IlaSerGlySerGlyGlnGlnIlnIlnValGlyMetCysIlnleuThrGlnAla	357		
466	CAGCAGCAGCAGCAGCAGCAGTTCACAGGCTCACAGCAGAGTGCATGCAGA	515		
358	ThrProIlaProSerGlnThrIlnIleSerSerAlaThrThrGlyThrGlnIlnI	374		

Percent Similarity: 49.922 Percent Identity: 27.786

alignment_block:

US-09-668-119-1 x US-09-041-886-23

Align seg 1/1 to: US-09-041-886-23 from: 1 to: 1185

```

163 CCTATGATGACATCCAGAGCCTGACTGGCGAGCCTGCTGGGAGCGGC 212
164 ProProProAlaLeuArgProLeuAsnAla...SerAlaSerPro 318
213 TGGAAATGGCATGCTCTCTGGGGCCGGGAGCAGCTCTGGGGGATGG 262
318 OGlyLeuGlyAlaGlnProLeu...ProGlyHis...Leu 330
263 GTAGCTTTGGCCATGAGACGCAATGCTCTGACAGGCGAGCCGCT 312
330 roserProTyrrAlaMetGlyGlnGlyMet...GlyGlyLeuPro 343
313 CTTGGG...ACCTCGGGATGGCCCTTCACAGATGGC 347
344 ProGlyProGlyGlyProGlyProGlyAlaProSerProHisSerLeu 360
348 TGTGCTGTCTACGGCACTCCACAGACCCAGCTGACAGCTCCAGAGTGG 397
360 oProAlaSerSerAlaPro...AlaPro...PrometArgPhe 373
398 CGCTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 447
374 ProTyrrSerSerSerSerSerSerSerAlaAlaAlaSerSer 388
448 GCGCTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 497
389 SerSerSerSerSerSerSerAlaSer...ProPheProAlaSer 402
498 GCAGAGTGC...CATGACAGCAGCTTCCAGAGCAGTGTGTC 535
403 GlnAlaLeuProSerTyrrProHisSerPheProProProHisSerLeu 419
536 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 585
419 rValSerAsnGlnProProLysTyrrThrGlnProSerLeuProSerGln 436
586 AATTGCAATCA... 596
436 laValTyrSerGlnGlyProProProProProTyrrGlyArgLeu 452
597 TCAAAATCAGACAGATACAGCAGCAGCAGCAGCAGCAGCAGCAG 643
453 AlaAsnSerAsnAlaHisProGlyProPheProProSerThrGlyAla 469
644 TACGACAGCTGCAAGTCCA...ACACAG 669
469 nSerThrAlaHisProProValSerThrHisHisHisHisGln 485
670 CAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 719
486 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisGlyAla 502
720 CCAGCCACCAATTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 762
502 rGlyProPro...ProProGlyAlaPheProHisProLeuGln 516
762 ... 762
516 lYgLYSerSerHisAlaHisProTyrrAlaMetSerProSerLeuGly 532
763 ...CCGCCCTCCAGAGCT...CTGCCAGCAGCAGTGTGCA 794
533 SerLeuArgProTyrrProProGlyProAlaHisLeuProProProHis 549
795 GCAGATGCATCACACAGCA...CCACAGCCGCGCACCA 831

```

```

549 rGlnValSerTyrrSerGlnAlaGlyProAsnGlyProProValSer 566
832 CAGCCCCAGCAGCAGCTCCAGTGTGTCAGAACCAACCATGACATCC 881
566 erSerAsnSerSerSerSerThrSerGlnGlySerTyrr...Pro 580
882 ACAGTCGACAGCAGCAGCAGCTTTGGTGTACAGGCGCAGACCTCC 931
581 SerHisProSerProSerGlnGlyProGlnGlyAlaProTyrrPhe 597
932 AATGCTGTATATCCCAACACCACTGAATTTGTCCGAGCTCCGATG 981
597 oProVal...ProHisValThrThrSerSerAlaThrLeuSerThr 612
982 GTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1031
612 AlAlaAlaThrValAlaSerSerPro...AlaGlyTyrrLysThr 625
1032 ACAGCAGCTGAGCTGCTGCCAGATGCTGGCTCCGGAGTCCAGCT 1081
626 AlaSerProProGlyProProProTyrrGlyLysArgAlaProSer 640
1082 AGAGCAGCCTCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1121
641 ...ProGlyAlaTyrrLysThrAlaThrProProGlyTyrrLys 654
1122 ...CGAGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1148
654 rGlySerProProSerPheArgThrGlyThrProProGlyTyrrArg 670
1149 CCGTCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1189
671 ThrSerProProAla...GlyProGlyThrPheLysProGlySerPro 687
1190 CACAGCCCACTCCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCT 1239
687 alGlyProGlyProLeuProProAlaGlyProSerGlyLeuProSer 702
1240 TTCCTGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1289
703 ...LeuProProProAlaAlaProAlaSerLysProProLeuSer 718
1290 GACC...CCACAGAACTTCAGTGTCC...TCAC 1318
718 aThrGlnHisLeuGlnGlnProAlaGlnGlnGlnGlnGlnGlnGln 735
1319 CT...GACCTTTAAACACACCTGTGAACCCAGCTGTGCATGAG 1362
735 roValProProAlaArgSerProSerProProProValVala 751
1363 CCAGCTGGCTCCAGCAG...GCGAGAGCAGCAGCAGCAGCAGCAG 1407
752 ProSerHisAlaSerGlnSerAlaArgPheAsnLysHisLeuAsp 768
1407 ... 1407
768 yPheAsnSerCysAlaArgSerAspLeuTyrrPheValProLeuGln 785
1408 ...CTGAAGCAGCTGTGCAAGTACATGACAGCCCTCGCCGCCAT 1448
785 erLysLeuAlaLysLysArgAlaAspLeuValGlnLysValAla 801
1449 ...GATCAACAGATGACAGCAGCAGCAGCAGCAGCAGCAGCAG 1495
802 AlaGlnGlnArgAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGln 818
1496 AGATGAAGACCTTCTGGACATTCAGACAGACCCCTGCAAGCGGTG 1545
818 gGlyLysGlnArgLysGlnGlnGlnGlnGlnGlnGlnGlnGln 833
1546 CTGAGACCTTGCAAAAGTGTGAGATCGCCCTGGAGCAAACTCAAG 1595
834 ValLysLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 845

```

1596 CATGGGTGC.....CCACTCCCCACGGCCCCGGTGCAC 1633
 846CysProSerLeuGlyProValProHisArgProPro...PheG 859
 1634 GACCAACACAGCTACCTATGCCAGCGCGCTCGATGCCGTCTGGCCA 1683
 859 LuProGlySerAlaValAlaThrValProProTyrLeuGlyProAsp 875
 1684 ACATCCCTGACCTGCTTCACACCATTCCTGTACCGACATTCGTCC 1732
 875 rProAlaLeuArgThrLeuSerGlyTyrAlaArgProHisValMetSer 891
 seq_name: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-991-300-2

seq_documentation block:

Sequence 2, Application US/08991300
 Patent No. 5973225
 GENERAL INFORMATION:
 APPLICANT: D'OVIDIO, RENATO
 APPLICANT: PORCEDDU, ENRICO
 APPLICANT: MERCHITELLI, CINZIA
 APPLICANT: CARDELLI, LUISA ERCOLI
 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
 NUMBER OF SEQUENCES: 6
 TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,300
 FILING DATE: 16-DEC-1997
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IT MI 96/A 002663
 FILING DATE: 19-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: OBLON, NORMAN F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 2264-0201-0X
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 369 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-991-300-2

alignment_scores:

Quality: 311.50 Length: 377
 Ratio: 1.639 Gaps: 16
 Percent Similarity: 50.398 Percent Identity: 32.095

alignment block:

US-09-668-119-1 x US-08-991-300-2 ..

Align seg 1/1 to: US-08-991-300-2 from: 1 to: 369

400 CTTCAGACAGCAGCAGCAGCAGCAGCTTCACAGCAGCAGCAGCGC 449
 6 LeuGluArgProSerGlnGlnGlnProLeuProProGlnGlnThrLeu 22
 450 GCTACAGCAGCAGCAGCAG.....CAGCAGCAGCAGCAGCAGTTCAGG 493
 22 rHisHisGlnGlnGlnProGlnGlnGlnProHisGlnPheProG 39
 494 CTTCAGCAGCTCCATGCAGCAGCAGCTTCAGCAGCTAGTCAGCAGCAG 543
 39 GlnGlnProCysSerGlnGlnGlnGlnProProLeuSerGln 55
 544 CAGCAG.....CTCCAGCAGCAGCAG.....CAGCAGCA 572
 56 GlnGlnProProPheSerGlnGlnGlnGlnProProPheSerGlnGln 72
 573 GCAGCATCTAATTAATTCATCATCA...AATCAGCAACAGATA.... 615
 72 ngInProValLeuProGlnGlnProSerPheSerGlnGlnLeuPro 89
 616CAGCAGCAGCAGCAGCAGCTCAGCAGTACAGCAGTGCAG 657
 89 rPheSerGlnGlnGlnProProPheSerGlnGlnGlnProVal 105
 658 CTCCAGCAGCAGCAGCAGCAGCAGCAGCAG.....CAGCA 698
 106 LeuProGlnGlnProSerGlnGlnGlnGlnGlnProProPheSerG 122
 699 GCAGCAGCAGCAGCTTGGAGCCCGCAGCCCAAT.....CAGCAGCC 742
 122 ngInLeuProProPheSerGlnGlnGlnProValLeuProGlnGln 139
 743 CGATCAGCAGCAGCAGCAGCTCCGCC...TCCAGCCTGCCCCAGCAG 789
 139 rPheSerGlnGlnGlnProProPheSerGlnGlnGlnProPhe 155
 790 CTGCAGCAGATGCATCAGCAGCAGCAGCCCGCAGCAGCCCA 839
 156 SerGlnGln.....GlnGlnProValLeuProG 165
 840 GCAGCTTCAGTTCCTCAGAACCAACCACTCCGCCAGCAGTGC 889
 165 ngInProProPheSerGlnGlnGlnGlnProLeuProGln.... 180
 890 AGACCCAGCTTGGTGTACAGCGCGCAAGCTCTCCCTGCAACAATGTTG 939
 181GlnProProPheSerGlnGln.....GlnProVal 191
 940 TATACCAACCACTGAAATTCGCCAGCTCCGATGGTGGAGCA 989
 192 LeuLeuGlnGlnGlnLeuProPheValHisProSerLeuLeuGln 208
 990 GCCCCAGTGCAGCCCGCAGTGCAGCAGCAGCAGCAGTTCAGACAG 1039
 208 uAsnProCysLysValPheLeuGlnGlnGlnCysSerProTrpAlaMet 225
 1040 CTCAGGCTGCCAGAGTGGCTCCCGAGTCCAGTCCAGCAGCAGCAGC 1089
 225 rGlnSer.....LeuAlaArgSerGlnMetLeuGlnGlnSer 238
 1090 CTCCCATGCTGTCCTGCGCGTCAACCGGCGCAGCAGTGCAGCCCGCA 1139
 239 CysHisValMetGlnGlnGlnCysGlnGlnLeuProGlnLeuProG 255
 1140 G..... 1140
 255 ngInSerArgTyrGlnAlaAlaLeuArgAlaLeuValTyrSerLeuLeu 272
 1141TCGATGCCCTCCCGCCAGCAGCTCC 1167
 272 GlnGlnGlnGlnGlnValGlnGlnGlnGlnGlnGlnGlnGlnPro 288
 1168 CCGCAGCCCGCCAGCCAGCTCAGCAGCC.....AATCCAGCAGTCA 1211


```
429 CCAGCAG.....CAGCAGAGCGCGCTACAGCAGCAGCAGCAGC 469
|||||
452 oGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 469
470 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 519
|||||
469 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 485
520 TTCCAGCAGTAGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 569
|||||
486 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 502
570 GCAGCAGCATCTAATTAAATTCATCAATCAATCAATCAATCAATCAAGC 619
|||||
502 nGlu.....ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 511
620 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 669
|||||
511 InGlnGlnGlnGln.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
670 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 719
|||||
521 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 537
720 CCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 766
|||||
537 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 554
767 CCTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 813
|||||
554 roGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 570
814 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 860
|||||
571 GlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 587
861 CCAACCATCAACATCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 910
|||||
587 gGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 604
911 AGGCGCAGCTCTCCGACAAATGTTATACCAACCAACCAACCAACCAAC 960
|||||
604 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 611
961 TTTGTCAGAGCTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1010
|||||
612 .....GlnAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 623
1011 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1060
|||||
623 pGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 637
1061 CTCGCCGAGTCCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1110
|||||
638 .....AspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 650
1111 TCACCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1160
|||||
651 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 667
1161 GCGGTCGCCGCGCAGCGCGC.....CAGCGCAGCTCAGCAGCAGCAGCAGC 1204
|||||
667 nAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 684
1205 AGCTGAGCTGTGCGCTGCGCCATCTCCAGTAGCTCTGCGCCAGCGCC 1254
|||||
684 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 700
1255 TCACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1304
|||||
701 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 716
```

```
1305 CAGTGTCCCTCAGCTGAGCTTTAAACACACCTGTGAACCCAGCTGTG 1354
716 .....
1355 TCATGAGCCAGCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1404
717 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 727
1405 AACCTAAGCAGCTGTGAGTAGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1454
728 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 743
1455 CAGATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1504
744 GlnAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 760
1505 GCCTTGTGACATTCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1554
760 uGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 773
1555 TTGCAAAAGTGTGAGATCGCCCTGAGAAACCTCAAGAA 1592
773 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 785
```

seq_name: /cgn2_6/plodata/2/iaa/5A_COMB.pep:US-08-209-747-2

seq_documentation_block:

Sequence 2, Application US/08209747

Patent No. 5733771

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V.

APPLICANT: Coldin, Mark

TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-3487

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/209,747

FILING DATE: 14-MAR-1994

CLASSIFICATION: 530

AUTHOR/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1447-104P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 832 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: Internal

ORGANISM: N. clavipes

TISSUE TYPE: minor ampullate gland

FEATURE:

NAME/KEY: CDS

LOCATION: 1..309

US-08-209-747-2

```
alignment_scores:
  Quality: 293.00      Length: 499
  Ratio: 1.118      Gaps: 17
Percent Similarity: 52.505      Percent Identity: 29.259
```

```
alignment_block:
US-09-668-119-1/rev x US-08-209-747-2 .
```

Align seg 1/1 to: US-08-209-747-2 from: 1 to: 832

[illegible]

```

1047 AGCCTAGCTGTCGTACTGCTGTGTGGTGCACCTCGGGGCTGA 998
324 .....ArglylaalaglyAlaaglYalaala 333
997 CTGGGGGCTGCTGCACACCATCGAGGCTCGGCAAAATTGAGTGCTG 948
      |||||:::||| ::|
334 AlaGlYAlaGlYAlaGlYAlaGlYTrgYl..... 344
947 TGGCATFACACAATTGTTCCAGGAGAGCTTGCGCTGTGACACCAAG 898
      ||| :::: |||
345 .GlyGIInserGIYTrgYAlaGlY.....A 353
897 CTGGGTCGTGCAGTTCGGGGAGTGTGATGGTGTCT..... 857
      |||||:::||| :|||
353 laGlYAlaAlaAlaAlaAlaAlaGlYAlaGlYglYAlaGlYglYTr 369
856 GAGCAACTGGAGGCTCGGGGCTGTGTGGCGGCTGTGCTGTGTG 807
      ::| :|||::||| :|||
370 GlYArGlYAlaGlYAlaGlYAlaGlYAlaAlaAlaGlYAlaGlY 386
806 TGATGCACTGCTGCGAGCGTGGGGCAGACGCTGGAGC..... 767
      |||||:::|||:::|||::||| :|||
386 yAlaAlaAlaGlYAlaGlYAlaGlYTrgYlYrGlYnglYnglYTrg 403
766 .....GCGAGGCTGTGCTGTGCATCGGTGCTGTA 731
      ||| |||||:::|||:::|||
403 lYAlaGlYAlaGlYAlaGlYAlaAlaAlaAlaGlYAlaGlY 419
730 TTGGTGGCTGGGCTCCAAAGCCTGCTGCTGCGTGCAGCGTGCCTG 681
420 GlYAlaGlYTrgYlYrGlYArGlYAlaGlYAlaGlYAlaAl 436
680 TGTGCGTGTGCT.....GTTGTGGAGCTGCAG 652
      |||::||| |||
436 acLYAlaGlYAlaGlYTrgYlYrGlYnglYnglYTrgYAlaGlY 453
651 CTGTGCTATTGCTGCAGCTGTGTGCTGCTGTATCTGTGTGT 602
453 laGlYAlaGlYAlaAlaAlaAlaAlaAlaThrGlYAlaGlYAla... 468
601 TTTGATGATGCAATTTATTAGATGCTGCTGCTGCTGCTGCTGG 552
469 .....GlyGlYTrgYlYrGlYAlaGlYAlaGl 478
551 AGCTGCTGCTGCTGCTGCACTACTGCTTGAACCTGCTGCTGCACT 502
      |||||:::|||:::|||
478 YAlaGlYAlaAlaAlaGlYAlaGlYAlaGlYTrgYlYrGlYAlaGlY 495
501 CTGCGAACCTGGAGAACGCTGCTGTGCTGCTGCTGCTGCTGCTGA 452
495 lYglYnglYnglYTrgYlYrGlYAlaGlYAlaGlYAlaAlaAla 511
451 GCGCCGCCGTGCTGCTCT.....GGAACCTGCTGT 417
      ||| :|||:::||| :|||
512 AlaaGlYAlaGlYAlaGlYAlaGlYTrgYlYrGlYAlaGlYAlaGl 528
416 TCCTGCTGCTGTGAGCGCACCTCGGAGCGAGCTGGGCTCT 371
      |||||:::|||:::|||
528 YAlaGlYAlaAlaAlaGlYAlaGlYAlaGlYAlaAlaAlaGlYAla 543
seq_name: /cgn2_6/plodata/2/1aa/5A_COMB.pep-US-08-458-298-2
seq_documentation_block:
? Sequence 2, Application US/08458298
? Patent No. 5756677
? GENERAL INFORMATION:
? APPLICANT: Lewis, Randolph V.
? APPLICANT: Colgin, Mark
? TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
? NUMBER OF SEQUENCES: 56
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Birch, Stewart, Kolasch & Birch

```



```

436 aglyalaglyalaglyalyltyrcltyrclgylglnlgltyrcltyrclalaglYa 453
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451 CTGCTGTAATTCGCTGCACAGCTGCTTTCCTGCTGCTGTAATCTTGGTGAT 602
452 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 laglyalaglyalalalalalalalalalalathrelyalaglyalglYa... 468
454
455
456
457
458
459
460 TTTGATGATGAATTTATTAAGATGCGTGCAGCGTCGCTGCTGCTGCTGCTGG 552
461 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
462
463
464
465
466
467
468
469 .....glycltyrcltyrclYargclYalaglyalagl 478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493 AGCTGCTGCTGCTGCTGCACACTACGCTTTGGACATGCTGCTGATGGCACT 502
494 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 YalagclYalalalalalalaglyalaglYalaglythnclgylYalaglyTYrc 495
496
497
498
499
500
501 CTGCTGAGCCTGGAACTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGA 452
502 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
503 YlglyglnlglYTYrclyalagclYalaglyalaglyalalalalalalal 511
504
505
506
507
508
509
510
511
512 GCGCGCGCTGCTGCGCGTCT .....GGAACCTGCTGCTGT 417
513 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514 AlagclYalaglyalaglyclYAlagclYTYrclyaragclYalaglyalagl 528
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529 TGTCTGCTGCTGCTGCAGCGCACCTGCTGAGCTGACGCTGAGTGTCT 371
530 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
531 YalagclYalalalalaglyalaglyalaglyalalalalaglyalalaglyal 543
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

seq.documentation.block:
Sequence 2, Application US/08853310
Patent No. 5948640
GENERAL INFORMATION:
APPLICANT: Randazzo, Filippo
TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppress
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/8553,310
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Guth, Joseph H.
REGISTRATION NUMBER: 31,261
REFERENCE/DOCKET NUMBER: 1228.003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3888
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

```

```
alignment_scores:
  quality: 292.50      Length: 543
  ratio: 1.312        Gaps: 22
Percent Similarity: 41.068  Percent Identity: 25.230
```

```
alignment_block:
US-09-668-119-1 x US-08-853-310-2    ..
Align seg 1/1  to: US-08-853-310-2  from: 1  to: 903
```

286	CAAATGCTCTCTCAGAGGCGACGCCCTCCCTGGACG.....	321
147	: : : : :	
147	ProMetSerAlaIaGlumIleGluValSerSerThrValIleThrAsnS	163
322TCGGGGATGGCCCCCTCCAGACATAGGGCTGTGTCTACGGCAA	364
163	rSerAsnSerAsnAspSerSerAsnAsnIleSerLeuCySerSerThrA	180
365	CTCCACAGACCCACAGCTGCAGCTCCAGAGGTGGCGCTGCAGCAGCAG	414
180	snSerLeuThrIleAsnGlnMetProHisGlnIleAlaSerGlnProGln	196
415	CAACAGCAGCAGCTTCCAGAGCAGCAGCAGCGG.....	447
197	GlnAsnAlaGlnSerAsnAlaGlnGlnIleArgIleIleuValAspS	213
448GCGCTACAGCAGCAGCAGCAGCAGCAGC	472
213	rAsnGlyIleIleIleGlyAsnPheLeuGlnGlnGlnIleArgIleGlnG	230
473	AGCAACAG.....CAGCATTTCCAGGCTCCAGCAGAGTGGCATGGAG	513
230	IleGlnGlnGlnIleuLeuGlnIlePheThrLeuGlnAlaIleAlaIleGln	246
514	CAGAGCTTCACAGCAGTGTAGTCACAGCAGCAGCAGCAGCTCCAGCAGCA	563
247	GlnGlnGlnGln.....GlnGlnGlnIleHisGlnGlnGlnGlnGlnGln	260
564	GCACAGCAGCAGCAG.....	576
260	ngIleGlnGlnGlnIleAlaThrSerSerAsnSerLeuGlyThrLeuProV	277
577CATCTAATTTAA	588
277	AlaIleuArgAsnGlyThrGlnGlnPheLeuSerProAsnLeuIleAla	293
589	TTGCATCAT.....CAAAATCAGCAGACAGATACGACGACAGCAACAGCA	632
294	GlnGlnIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisGlnGlnGln	310
633	GCTG.....CAGCGAATACACAGCAGTCGCTCCAC	664
310	AlaIleThrAlaGlnGlnIleGlyHisGlnIleIleGlnIlePheAlaLeuGlnG	327
665	AACACAGCAGCAGCAGCAGCAGCAG.....	687
327	IleAlaGlnIleuHisGlnIleArgIleuLeuAlaGlnAlaIleAlaAsnAsn	343
688CAGCAGCAGCAGCAGCAGCAGCTTTGGAGGCCACGACCAACAT	731
344	LeuIleGlnGlnGlnGlnGlnGlnGlnGlnGlnIleAsnValAlaLeuProThrTh	360
732	TCAGCAGCCACGAGTCAGCAGCAGCA.....CAGCTCCGC	766
360	rGlnAlaIleAspHeIleAlaIleAspProLeuAsnIleIleSerMetThrArgP	377
767	CTCCACAGGCTGTGCCCGCAGACGCTGCAGCAGATGCATCAGACA.....	810
377	rAlaAsnAlaSerProThrThrAlaAlaThrThrAlaAsnThrAlaSer	393
811CAGCA	815
394	IleProSerAlaIleAlaAsnValAlaValAlaThrGlyAlaGlnIleGln	410
816	CCACAGCGCGGCA...CCACAGCGCCAGCAGCTCCAGTGTGTCAACAC	862
410	ngIleSerProIleValProAlaIleProGlnGlnIleThrValGlnGlnGlnG	427

```

863 AACCATGACAACTCCGCGACAGTCCGACGCGCCGCTTGTGTGACAG 912
    ||| ||| ||| |||
427 LLeuAlaAsnHisAsnSerAsnMetGlnGlnProAsnValLeuThr 443
    ||| ||| ||| |||
913 GCGAAGCTCTCCCTGGACAAATGTGTATACCAACCACTGAATTT 962
    ||| ||| ||| |||
444 MetLysThrLeuProSer..... 450
963 TGTCCGAGCTCCGATGTGTCGACGAGCCGCGACGCGCGAGTGC 1012
    ||| ||| ||| |||
451 ...GlyValProThrThrIleAlaGlnGlnAlaGlnProLysMet. 465
1013 AGCAGCAGCAGCAGCAGTACAGACAGCTGCGCGAGTGTGCGCT 1062
    ||| ||| ||| |||
466 .....ProThrGlyLysGlyArgLysAlaThrSerAsnArgLeuPro 479
1063 CCGCGAGTC.....CAGGTACGCGACGACGACCT 1091
    ||| ||| ||| |||
480 ProGlyAlaValAsnLeuGlnArgThrTyrGlnIleCysGlnAlaVal 496
1092 CCGCATGCTGCTCCGCTGACGCGGCGCAGAGTGCAGACCGCGAGT 1141
    ||| ||| ||| |||
496 e.....GlnAsnSerProAsnArgLysAlaGln 509
1142 CGATGCGCGCTCCGCG.....CAGCG..... 1164
    ||| ||| ||| |||
509 euArgProAlaAlaIleLeuAsnGlnIleGlnProThrThrThr 525
1165 TCCCGCGAGCGCGCGCGCGCG.....AGCTC 1190
    ||| ||| ||| |||
526 AlaProAlaProIleAsnProValThrLeuAsnValSerThrValAla 542
1191 ACAGCCCAACTCCAGCTCAGCTGTGC..... 1218
    ||| ||| ||| |||
542 aThrProMetSerAsnIleThrThrAlaThrGlySerMetAlaAlaVal 559
1218 ..... 1218
559 AlaAlaAlaProGlnAsnValLeuLysGlnGlnLeuLeuVal 575
1219 .....CCTGCGCGACCTCCGCGAGTACGCT 1241
    ||| ||| ||| |||
576 SerGlyAlaValGlyAlaLeuProAlaGlyLeuProProAsnVal 592
1242 CCG..... 1245
592 LMetGlyAlaGlyArgProGlyValTyrLysValIleGlyProArgMet 609
1246 .....CCGACGCG..... 1254
609 eGlyPheArgArgLysTyrValGlnArgLysProSerProThrThr 625
1255 .....TCACCGAGCGCTCCGCGACGCGCGCGAGCG 1286
626 LeuIleArgHisValPheSerProGlyGlyAlaThrAlaThrAla 642
1287 GCGGAGCGCGACG.....ACCTTCAGCTCCGCTGACGCTG 1321
    ||| ||| ||| |||
642 aGlnIleLeuGlnMetLeuGlnGlnIleHisIleSerThrThrSerPro 659
1322 GAGCTTAAACACAGCTGTGAACCGCG 1350
659 aLProValGlnAsnProGlnGlnProAla 668
seq_name: /cgn2_6/ptoddata/2/1aa/5A_COMB pep: US-08-415-751-35
seq_documentation_block:
; Sequence 35, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.

```

```

APPLICANT: GUT, JIRI
TITLE OF INVENTION: POLYPEPTIDES BINDING ANT-
TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: PHILLIPS, MOORE, LEMPTO & FINLEY
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071,880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/891,301
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolezalova
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.19-2 (HND)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
FEATURE:
NAME/KEY: Positions coded by nonsense codons are
NAME/KEY: Identified as Xaa.
US-08-415-751-35

alignment_scores:
Quality: 291.50 Length: 162
Ratio: 3.512 Gaps: 12
Percent Similarity: 51.235 Percent Identity: 36.420

alignment_block:
US-09-668-119-1/rev x US-08-415-751-35 ..
Align seg 1/1 to: US-08-415-751-35 from: 1 to: 362
717 CTCGAAAGCTGCTGCTG.....CTGCTGCTGCTGCTGCTGCT 677
    ||| ||| ||| |||
78 LeuGlnAsnArgLeuValGlu*****LeuGlnLeuArgLeuValVa 94
    ||| ||| ||| |||
676 GCTGTGCTGTGTGTGAGCTGACGCTGCTATTGCTGACG..... 633
    ||| ||| ||| |||
94 LLeuLeu*****LeuGlnPheAspLeuGlyValLeu*****Trp 110
    ||| ||| ||| |||
632 .....TGCTGTGCTGCTGCTATC.....TGTTG 607
    ||| ||| ||| |||
111 ***ArgLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 127

```



```

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-089-862-11
seq_documentation_block:
  Sequence 11, Application US/08089862
  Patent No. 5723588
  GENERAL INFORMATION:
  APPLICANT: Donofrio, David A
  APPLICANT: Stedronsky, Erwin R
  TITLE OF INVENTION: PROTEIN-ENRICHED THERMOPLASTICS
  NUMBER OF SEQUENCES: 12
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
  STREET: 4 Embarcadero Center, Suite 3400
  CITY: San Francisco
  STATE: California
  COUNTRY: USA
  ZIP: 94111-4187
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patent Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/089,862
  FILING DATE: 09-JUL-1993
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  NAME: Rowland, Bertam I
  REGISTRATION NUMBER: 20,015
  REFERENCE/DOCKET NUMBER: A-58854/BIR
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 781-1989
  TELEFAX: (415) 398-3249
  TELEX: 910 277299
  INFORMATION FOR SEQ ID NO: 11:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1018 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: peptide
  US-08-089-862-11

alignment_scores:
  Quality: 287.00      Length: 532
  Ratio: 0.966        Gaps: 18
  Percent Similarity: 55.827    Percent Identity: 27.820

alignment_block:
  US-09-668-119-1/rev x US-08-089-862-11 ..
  Align seg 1/1 to: US-08-089-862-11 from: 1 to: 1018

1719 GGTACGAGGATGTTGAAGACAGGTGAGCGATGTGGCCAGCAGCA 1670
|||||
102 GYAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGly 118
1669 TCCAGGAGG.....GCTGGCATAGTACTCTCTGTTGGTGGTGG 1629
|||||
118 YAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 134
1628 CACCGGGGCGGTGGGAGTGGGACGCGCATGTGATTTGAGTTTCG 1579
|||||
134 YSerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGly 151
1578 CAGGCGCATCTCACACTTTTGCAGCTTTCAGGGAGCGCTTCGAGG 1529
|||||
151 TGGYAlaSerIleYsValAlaValSerAlaGlyProSerAlaGlyTyr 167
1528 GGTCTG.....TCAGATGTCGAGAAAGGCTTTCATCTTACTCAGG 1488
|||||
168 GYAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 184

```

```

1487 TCCTTTTCTGCTCTGTTCTGTCGATCTTGTGATCATCGCGCAG 1438
184 Y.....SerGlyAlaGly 188
1437 GGGCTCGATGTAATTCGACAGCTGCTCA...GCTTTCGACGTAATCT 1391
|||||
188 YAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGly 204
1390 GCTCTCAGCTGCTGCGTGGACCGACGCTGAGCTGATCAGAGCTGGTTC 1341
|||||
205 AlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 221
1340 ACAGTGTCTTAAAGCTCCAGCTGAGGAGGACCTGAAGTCTGTGGGT 1291
|||||
221 TGYAlaAlaProGlyAlaSerIleYs.....ValAlaValSerA 235
1290 CCGCGCCGTCAGCTGCTCTGTCGAGGCGTGGCGGTGAGGCG...TGGCA 1244
|||||
235 IacGlyProSerAlaGlyTyrGlyAlaGlyAlaGlySerGlyAlaGlyAla 251
1243 GGAACCTACTGGAGATGGGCGAGGCCAGAGCTGACGTTGAGTTGGGC 1194
|||||
252 GlySerGlyAlaGlyAlaGlySerGly..... 260
1193 TGTGAGCTGGGCTGGCGGCTGGCGGAGCGGCTGGGAGGAGGCGAT 1144
|||||
261 ...AlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaG 276
1143 CGACTGCGGGGCTGTCACACTGCTGCGTGCACCTGGGAGGAGCAGCTGG 1094
|||||
276 YSerGlyAlaGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGly 292
1093 GGAGCTGCTGCTGCTGACCTGAGCTCGGAGCGCACCATCTGGGCGCC 1044
|||||
293 AlacGlySerGlyAlaAlaProGly..... 300
1043 TGAGCTGCTGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 994
|||||
301 ...AlaSerIleYsValAlaValSerAlaGlyProSerAlaGlyTyrG 316
993 GCGCTGCTGACACCAATCGAGCTGGACAAATTCAGTGGTGGG 944
|||||
316 YAlaGlyAla..... 319
943 TATACACATTTTCAGGAGGAGCTGGCGCTGCTGTCACACAAAGCTGG 894
|||||
320 .....GlySerGlyAlaGlyAlaGlySerGlyAlaGly 330
893 GTCGCGACTGTGGCGGAGT.....GTGATGTTGTTCTGAGC 853
|||||
330 YAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGly 347
852 AACTGAGGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 803
|||||
347 IacGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySer 363
802 GCATCTGCTCAGCTGCTGGGCGAGCGTGGAGGCGGAGGCTGTGGC 753
|||||
364 GYAlaGlyAlaGlySerGlyAlaAlaProGly..... 374
752 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 703
|||||
375 ...AlaSerIleYsValAlaValSerAlaGlyProSerAlaGlyTyrG 390
702 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 653
|||||
390 YAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 406
652 GCTGTCTATTCGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
|||||
407 SerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly..... 418

```



```
301 ...AlaserIleYsValAlaValSerAlaGlyProSerAlaGlyTyrG 316
993 GGGCTGCTGCACACCATCGAGCTCGACAAATTTCAGTGTGTTGGC 944
316 lYAlaGlyAla..... 319
943 TATACACATTTCACAGAGAGCTTGCCTGTGACACCAAGCTGG 894
320 .....GlySerGlyAlaGlySerGlyAlaGlySerGlyAlaGly 330
893 GTCCTCCGACTGTGGCGGAGTT.....GTGATGCTGTGTTGAGC 853
330 YAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlySerGlyAla 347
852 AACTGAGAGCTGTGGGCTGTGGGCTGTGGGCTGTGGTGTGAT 803
347 lAglyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySer 363
802 GCATCTGCTGAGCTGTGGGCGAGACCTGGAGGCGGAGGCTGTGGC 753
364 GlyAlaGlyAlaGlySerGlyAlaAlaProGly..... 374
752 TGCTGCATCGGTGGCTGTGATGTGGCTGTGGCTTCCAAAGCTGTCTG 703
375 ...AlaserIleYsValAlaValSerAlaGlyProSerAlaGlyTyrG 390
702 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 653
390 lYAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 406
652 GCTGTGCTATTCGCTGACGCTGCTGTGCTGCTGCTGCTGCTGCTGA 603
407 SerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly..... 418
602 TTTGATGATGCAATTAAATTAGATGCTGCTGCTGCTGCTGCTGCTG 553
419 .....SerGlyAlaGlyAlaGlySerGlyAlaGly 428
552 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAC 503
428 lYAlaGlySerGlyAlaGlyAlaGlySerGly...AlaGlyAlaGlySer 443
502 TCTGCTGAGCTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453
444 GlYAlaAlaProGlyAlaSerIleYsValAlaValSerAlaGlyProSe 460
452 AGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
460 rAlaGlyTyrGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyA 477
402 CAGCGCCACTGCTGTGAGCTGACGTGGTCTGTGAGATTGCCGTAGACA 353
477 lAglyAlaGlySerGlyAlaGlyAlaGlySer...GlyAlaGlyAlaGlys 493
352 CGACAGCCATGCTGTGAGGGGCGCATCCCGAGTCCCGAGAGCGGCTGC 303
493 erGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAla 509
302 CTTGAG...AGAGCATTTGGCTGTCCCATGGCACCAAGCTACCATCCC 256
510 GlySerGlyAlaGlyAlaGlySerGlyAlaAlaProGlyAlaSerIleY 526
255 GCCCAAGAGACTGTCCGGGCGCCGAGAGGAGCATGCCAATTCACGCGCTC 206
526 sValAlaValSerAlaGlyProSerAlaGlyTyrGlyAlaGlyAlaGlys 543
205 CCGCAGCAGGTCCGCGCATGAGGCTGTGAGTGCATTC 167
543 erGlyAlaGly...AlaGlySerGlyAlaGlyAlaGlySer 555
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2002, 17:49:47 ; Search time 86.87 Seconds
(Without alignments)
4536.338 Million cell updates/sec

Title: US-09-668-119-1

Perfect score: 1740

Sequence: 1 atgaggaagaagcgtgtgtgc.....cacatcttcacgcatga 1740

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCtus.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181.2	10.4	3489	2	US-08-728-323A-1
2	181.2	10.4	32207	2	US-08-770-379-20
3	181.2	10.4	32207	2	US-08-757-669A-20
4	171.8	9.9	543	6	5273901-6
5	170	9.8	533	6	5482709-5
6	142.6	8.2	397	3	US-09-253-691-3
7	130.8	7.5	234	1	US-08-469-802B-3
8	130.8	7.5	234	2	US-08-267-803B-3
9	130	7.5	688	4	US-08-998-416-915
10	126.6	7.3	6530	2	US-08-146-930-1
11	126.6	7.3	6530	2	US-08-458-240-1
12	126.6	7.3	6530	5	PCt-US93-03993-1
13	120.6	6.9	477	4	US-09-135-994-1
14	118.8	6.8	2580	2	US-09-050-863-2
15	118.8	6.8	5452	2	US-09-130-114-1
16	118.8	6.8	9600	4	US-08-910-647-1
17	118.8	6.8	10596	1	US-07-884-811-15
18	118.8	6.8	10596	1	US-07-885-971-15
19	118.8	6.8	10596	1	US-08-087-783A-15
20	118.8	6.8	10596	1	US-08-194-088B-15
21	118.8	6.8	10596	2	US-08-194-087-15
22	118.8	6.8	10596	5	PCt-US93-04648-15
23	114.8	6.6	2793	1	US-08-209-747-1
24	114.8	6.6	2793	1	US-08-458-298-1
25	113.8	6.5	2214	4	US-09-043-303-7
26	113.4	6.5	2214	3	US-08-864-038A-1
27	113.4	6.5	3331	3	US-08-864-038A-2

C	28	113.4	6.5	3331	3	US-08-864-038A-4	Sequence 4, Appli
	29	111.2	6.4	195	1	US-08-469-802B-2	Sequence 2, Appli
	30	111.2	6.4	195	2	US-08-267-803B-2	Sequence 2, Appli
C	31	110.8	6.4	786	1	US-08-403-379A-2	Sequence 2, Appli
	32	110.8	6.4	786	2	US-08-929-414-2	Sequence 2, Appli
C	33	110.8	6.4	9551	1	US-08-056-200-93	Sequence 93, Appl
	34	110.8	6.4	9551	2	US-08-800-644-93	Sequence 19, Appl
	35	109	6.3	2190	4	US-09-625-188-19	Sequence 17, Appl
	36	105.8	6.1	165	4	US-09-103-429A-17	Sequence 1, Appli
	37	101.8	5.9	2455	4	US-08-781-891-208	Sequence 208, App
C	38	101.8	5.9	16442	3	US-08-425-069-3	Sequence 3, Appli
	39	98.8	5.7	1995	1	US-08-317-844B-3	Sequence 3, Appli
	40	98.8	5.7	1995	2	US-08-103-429A-2	Sequence 2, Appli
	41	97.4	5.6	2821	4	US-08-469-802B-6	Sequence 6, Appli
	42	96	5.5	154	1	US-08-267-803B-6	Sequence 6, Appli
	43	96	5.5	154	2	US-07-814-220-3	Sequence 3, Appli
	44	93.6	5.4	336	2	US-07-814-220-4	Sequence 4, Appli
C	45	93.6	5.4	336	2	US-07-814-220-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5946676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
; US-08-728-323A-1


```

RESULT 12
PCT-US93-03993-1/C
; Sequence 1, Application PC/TUS9303993
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Roehrig, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: CONSTITUTIVE AND INDUCIBLE EPIDERMAL
; TITLE OF INVENTION: VECTOR SYSTEMS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19930428
; PCT/US93/03993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6530 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US93-03993-1

```

```

Query Match
Best Local Similarity 7.3%; Score 126.6; DB 5; Length 6530;
Matches 407; Conservative 0; Mismatches 424; Indels 6; Gaps 2;
QY 360 ggaactccacagacacagctgacgtccagcaggtgagctgacagcagcagcaaca 419
DB 3635 gccgctccgagcgtgagacgtccgcccgcagagagagccgacgagcagcagc 3576
QY 420 gcaagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagc 479
DB 3575 gccgctccgagcagcagcagcagcagcagcagcagcagcagcagcagc 3516
QY 480 gcaagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagc 539
DB 3515 gcaagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagc 3456
QY 540 gcaagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagc 599
DB 3455 gccgctccgagcagcagcagcagcagcagcagcagcagcagcagcagc 3396
QY 600 aatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 659
DB 3395 gtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3336
QY 660 ccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 719
DB 3335 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3276

```

```

QY 720 ccagcacaacatccagcagcagcagcagcagcagcagcagcagcagcagcagc 779
DB 3275 accgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3216
QY 780 gccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 839
DB 3215 gccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3156
QY 840 gccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 899
DB 3155 gccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3095
QY 900 ttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 959
DB 3096 --tgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3039
QY 960 attgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1019
DB 3038 gtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2979
QY 1020 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1079
DB 2978 gccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2919
QY 1080 ccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1139
DB 2918 cttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2862
QY 1140 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1196
DB 2861 accagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2805

```

```

RESULT 13
US-09-135-994-1
; Sequence 1, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Rannum et al.
; TITLE OF INVENTION: SCAT GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; EARLIER FILING DATE: 1997-08-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-135-994-1
Query Match
Best Local Similarity 6.9%; Score 120.6; DB 4; Length 477;
Matches 183; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 468 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 527
DB 51 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 110
QY 528 agtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 587
DB 111 agtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 170
QY 588 attgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 647
DB 171 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 230
QY 648 acagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 707

```


Fri Mar 1 09:11:59 2002

us-09-668-119-1.rni

Page 10

Search completed: February 28, 2002, 19:18:22
Job time: 5315 sec

2 . . .

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2002, 17:59:17 ; Search time 171.43 Seconds
(without alignments)
8701.779 Million cell updates/sec

Title: US-09-668-119-1

Perfect score: 1740
Sequence: 1 atgagagaagctggtgtgctc.....cacatcttcacagcatga 1740

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 segs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_1101:*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1725.8	99.2	2556	22	AA159455
2	1647.2	94.7	3438	22	AA161241
3	939.2	54.0	1388	21	AA659039
4	338.6	19.5	386	22	AA65986
5	241.4	13.9	316	22	AA120017
6	241.4	13.9	316	22	AA145215
7	241.4	13.9	316	22	AA105725
8	241.4	13.9	436	22	AA110775
9	241.4	13.9	436	22	AA132033
10	241.4	13.9	436	22	AA100701
11	191.6	11.0	486	22	AA75507

12	181.2	10.4	3489	21	AAA30290
13	181.2	10.4	3489	22	AA682901
14	181.2	10.4	32207	20	AAV73805
15	181.2	10.4	137507	19	AAV19941
16	171.8	9.9	543	13	AAQ23092
17	167.4	9.6	5120	22	AA684677
18	166.4	9.6	14704	13	AAQ20685
19	163.2	9.4	10268	17	AA133007
20	156.8	9.0	567	21	AA29550
21	153.6	8.8	658	21	AA245602
22	149.2	8.6	1159	21	AA659240
23	147.6	8.5	1037	21	AA659242
24	147.6	8.5	1472	21	AA659241
25	140	8.1	397	20	AA689891
26	140	8.0	6794	21	AA455597
27	131.8	7.6	8201	21	AA688864
28	131.6	7.6	629	21	AA708420
29	131.6	7.6	3015	20	AA632027
30	131.6	7.6	3015	22	AA690084
31	130.8	7.5	234	16	AAQ84832
32	126.6	7.3	6530	14	AAQ51557
33	126.6	7.3	6530	20	AA622072
34	126.6	7.3	6530	22	AA68948
35	125	7.2	6463	21	AA645239
36	121	7.0	422	21	AA679295
37	120.6	6.9	477	21	AA445437
38	120.4	6.9	799	19	AA655831
39	120.4	6.9	799	19	AA655831
40	118.8	6.8	1925	20	AA690924
41	118.8	6.8	1926	21	AA650254
42	118.8	6.8	1926	22	AA682902
43	118.8	6.8	2580	21	AA675454
44	118.8	6.8	5452	20	AA690923
45	118.8	6.8	8705	20	AA623778

ALIGNMENTS

RESULT 1	
AA159455	AA159455 standard; cDNA: 2556 BP.
XX	XX
AC	AA159455;
XX	XX
DI	22-OCT-2001 (first entry)
XX	XX
DE	Human polynucleotide SEQ ID NO 1658.
XX	XX
KW	Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200153312-A1.
XX	XX
PD	26-JUL-2001.
XX	XX
PF	26-DEC-2000; 2000WO-US34263.
XX	XX
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0586042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	XX

Kaposi's sarcoma-a
Nucleotide sequenc
KSHV LTR DNA (nucl
KSHV long unique c
Antigen tc-7a gene
DNA sequence of hu
PKS 741 insert con
Mouse SRY-related
HIV codon altered
cDNA sequence of a
An EcoRI fragment
Exons E, C and A o
Exons D, C, B and
Spinocebellar at
cDNA sequence of h
Human dentin stalo
Fusarium venenatum
Human MERT1 relate
246970 cDNA clone.
Spinocebellar at
Loricrin gene. Ho
Nucleotide sequenc
Mouse loricrin gen
DNA encoding angio
Human lung tumour-
Probe #15232 used
Nucleotide sequenc
Epstein Barr Virus
Epstein Barr Virus
EBV tethering prot
Nucleotide sequenc
Anti-sense strand
Vector pshuttle DN


```

QY 301 gggcagccgctctcggagacctcgggagatggccctcagcagatgctgtctctacg 360
    |||
DB 398 gggcagccgctctcggagacctcgggagatggccctcagcagatgctgtctctacg 457
QY 361 ggaattccagagaccagctgtcagctcagcagatggtggctgagcagcagcagcagcag 420
    |||
DB 458 ggaattccagagaccagctgtcagctcagcagatggtggctgagcagcagcagcagcag 517
QY 421 cagcagttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 480
    |||
DB 518 cagcagttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 577
QY 481 cagcagttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 540
    |||
DB 578 cagcagttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 637
QY 541 cagcagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
    |||
DB 638 cagcagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 697
QY 601 aatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
    |||
DB 698 aatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 757
QY 661 caacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
    |||
DB 758 caacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 817
QY 721 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 780
    |||
DB 818 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 877
QY 781 ccccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 840
    |||
DB 878 ccccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 937
QY 841 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900
    |||
DB 938 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 997
QY 901 ttgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 960
    |||
DB 998 ttgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1057
QY 961 ttgtgtc 966
    |||
DB 1058 atcattc 1063

```

RESULT 4

AAF65986/c
ID AAF65986 standard; cDNA: 386 BP.

AC AAF65986;

XX 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1742.

XX Human; cytosolic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

XX (CHIR) CHIRON CORP.
PA (HSE-) HYSED INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Ckenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Grain B;
XX WPI: 2001-091805/10.

Library of polynucleotides for diagnosing a cancerous state of a
mammalian cell and detecting cancer, particularly of the colon or
prostate, comprises 3351 human polynucleotide sequences -
Claim 9; Page 793; 1046pp; English.

The present sequence is one of 3351 sequences in a library of human
polynucleotides. The library is used to detect differentially expressed
genes correlated with a cancerous state of a mammalian cell and can
detect colon, prostate, breast and lung cancer. The library can be used
to produce probes for detection of mRNA and to produce additional copies
of the polynucleotides. The probes can be used for chromosome mapping of
the polynucleotide and for detection of transcription levels. Ribozymes
or antisense oligonucleotides can be generated. The polynucleotides and
their gene products are used as genetic or biochemical markers (e.g. in
blood or tissues) that will detect the earliest changes along the
carcinogenesis pathway and/or monitor the efficacy of therapies and
preventive interventions. The polynucleotides, polypeptides and
antibodies against them can be used in pharmaceutical compositions to
treat the cancers and proliferative disorders such as neoplasia,
dysplasia and hyperplasia.

Sequence 386 BP; 50 A; 88 C; 138 G; 110 T; 0 other;

Query Match 19.5%; Score 338.6; DB 22; Length 386;
Best Local Similarity 96.0%; Pred. No. 8.4e-51;
Matches 358; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```

QY 490 caggttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 549
    |||
DB 386 CATGCTCAGCAGAGTG-CATGCGAGCAGAGTTCTTCTAGCAGTAGTCAGCAGCAGCAGTAG 328
    |||
QY 550 ctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 609
    |||
DB 327 CTCAGCTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 268
    |||
QY 610 cagatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 669
    |||
DB 267 CAGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 208
    |||
QY 670 caacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 729
    |||
DB 207 CAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 148
    |||
QY 730 attcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 789
    |||
DB 147 ATTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 88
    |||
QY 790 ctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 849
    |||
DB 87 CTGCAGCAGATGATTCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 28
    |||
QY 850 gtgtcagcagcagc 862
    |||
DB 27 GTTGTCTCAGAAC 15

```

RESULT 5

AAI20017
ID AAI20017 standard; DNA: 316 BP.

Ratio: 1.587 Gaps: 30
Percent Similarity: 43.586 Percent Identity: 29.675

alignment_block:

US-09-668-119-1 x T03454 ..

Align seg 1/1 to: T03454 from: 1 to: 5262

```

1226 CCGCCCTGGGGGGGGGGGACAGCTCTGGGGGGGATGGTAGCTTTGGTGC 275
1227 |||||||:|||||:|||||:
3519 ProtrGlnGlyProGlnGlyMetLeuGly..... 3528
276 CATGGACAGCCAAATGCTCTCAGAGGAGCCGCTCTGGAGCTCGG 325
277 :|||:|||||:|||||:|||||:
3529 .ProAlaGlnValAlaValLeuGlnGlnGlnHisProGlyAlaLeuGlyP 3545
326 GGATGGCCCCCTCAGACATGGCTGTGTCTAAGGCAATCCAGACAGC 375
3545 roGlnGlyProHisArgGlnValLeuMetThrGlnSerArgValLeuSer 3561
376 CAGCTGAGCTCCAGCAG.....GT 395
3562 SerProGlnLeuAlaGlnGlnGlnGlyLeuMetGlyHisArgLeuVal 3578
396 GCGGCTGACAGCAGCAGCAGCAGCAGCAGTTCAGCAGCAGCAGCAGG 445
3578 lThrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlySerH 3595
446 CCGCG..... 450
3595 eTrAlaGlyLeuSerHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 3611
451 .....CTACAGCAGCA 461
3612 ProGlyLeuSerAlaGlnProMetGlySerLeuGlnGlnGlnGlnGln 3628
462 GCAGCAGCAGCAGCAGCAGCAGCAGTTCAGCAGCAGTCCAGAGTGCATGC 511
3628 ngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3642
512 AGCAGAGTTCAGCAGCAGTGTGCAG...CAGCAGCAGCAGCAGTCCAGCAG 558
3642 lngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3658
559 CAGCAGCAGCAGCAGCAGCAGTCAATTAATTCATCATCAAAATAGCA 608
3659 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3671
609 ACAGATACAGCAGCAGCAGCAGCAGTGCAGCAATAGCAGCAGCTGCAGC 658
3671 ngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3688
659 TCCACACAGCAGCAGCAGCAGCAGCAG..... 690
3688 lngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3704
691 .....CAGCAGCAGCAGCAGCAGCTTGGCA 716
3705 GlnSerArgThrLeuLeuSerProGlnGlnGlnGlnGlnGlnGlnGlnGln 3721
717 GCGCCAGCAGCAATTCAGCAGCAGCAGTGCAG..... 750
3721 AlauGlyProGlyMetProAlaGlyProLeuGlnHisPheSerSerProG 3738
750 ..... 750
3738 lylAlaLeuGlyProThrLeuLeuThrGlyGlyGlnGlnAsnThrVal 3754
750 ..... 750
3755 AspProAlaValSerSerGlnAlaThrGlnGlyProSerThrHisGlnGln 3771
750 ..... 750

```

```

3771 yGlyProLeuAlaIleGlyThrThrProGlnSerMetAlaThrGlnProG 3788
750 ..... 750
3788 lylGlnValLysProSerLeuSerGlyAspSerGlnLeuLeuValGln 3804
751 ...CAGCCAGAGCCCTCCGCTCCAG.....GC 776
3805 ProGlnProGlnProGlnProSerSerLeuGlnLeuGlnProProLeuAl 3821
777 TCTGCCCCAGCAGCTGCAGCAG.....ATGCAT..... 804
3821 GLeuProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3838
804 ..... 804
3838 lylGlySerHisGlyGlnLeuGlySerGlySerSerGlnAlaSerSer 3854
804 ..... 804
3855 ValProHisLeuLeuAlaGlnProSerValSerLeuGlyAspGlnProGln 3871
805 ...CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 850
3871 ySerMetThrGlnAsnLeuLeuGlyProGlnGlnProMetLeuGlnArg 3888
851 TTGCTCAGAACCCAACTCACAACCTCCGCGCAGCAGTGCAGACCCAGCT 900
3888 rMetGlnAsnAsnThrGlyProGlnProPro.....LysProGlyPro 3902
901 TTGCTGTCCAGAGCCAGCTCCCTGGACAATGTTGTATACCAGC 950
3903 ValLeuGlnSerGlyGlnGlyLeuProGly.....ValGlyIleMetPr 3917
951 ACCACTGAATTTGTCCAGCTCCGATG.....GTGGTCAGCAGCAGCC 994
3917 oThrValGlyGlnLeuArgAlaGlnLeuGlnGlyValLeuAlaLysAsnPr 3934
995 CA.....GTGCAGCCCCAGTGCAGCAGCAG..... 1020
3934 roGlnLeuArgHisLeuSerProGlnGlnGlnGlnGlnGlnGlnGlnGln 3950
1021 .....CAGCAGCAGCTACAGCAGCTCAGCTCCGATGATG..... 1059
3951 LeuMetGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3967
1060 .....GCTCCGAGATCCAGTGCAGC.....CAGAGCAGCTC..... 1092
3967 oThrGlnGlnProGlyThrGlnThrGlnThrProLeuGlnGlyLeuGlnGlyC 3984
1093 .....CCCATGCTGTCTCCGCGGTACCCGGGSCAGAGTGCAGACCCCG 1137
3984 ySglnProGlnLeuGlyGly...PheProGlyProGlnThrGlyProLeu 3999
1138 CAGTGCAGTCCGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1187
4000 GlnGlnLeuGlyAlaGlyProArgProGlnGlyProProArgLeuProAl 4016
1188 TCCACAGCCCAATCCACAGTGCAGCTGCAG..... 1218
4016 a.....ProProGlyAlaLeuSerThrGlyProValLeuGlyProValH 4031
1219 ..CTGCCCCCATCTCCAGTACGCTCTCCGCGCAGCC...CTCAGCCGAG 1263
4031 lSProThrProProProSerSer...ProGlnGlnProLysArgProSerG 4047
1264 CCGTCCAGAGCCAGCAGTGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1313
4047 lneupProSerProSerSerGlnLeu...ProThrGlnAlaGlnLeuPro 4062
1314 CTCACCTGAGCTTTAAACACACCTGTGACCCGAGCTGTCTGACGACC 1363
||| ||| :||| |||

```


C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
 C:Accession: A38660; B38:46
 R:Wood, L.; Mills, M.; Hetzenbuehler, N.; Vogell, G.
 J. Biol. Chem. 266, 4024, 1991
 A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin
 A:Reference number: A38660; MUID:91154184
 A:Accession: A38660
 A:Molecule type: DNA
 A:Residues: 1-223 <W02>
 A:Cross-references: GB:M7760; NID:g200963; PIDN:AAA0107.1; PID:g200964
 A:Note: This is a correction
 R:Wood, L.; Mills, M.; Hetzenbuehler, N.; Vogell, G.
 J. Biol. Chem. 265, 21375-21380, 1990
 A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin
 A:Reference number: A38346; MUID:91065960
 A:Accession: B38346
 A:Molecule type: DNA
 A:Residues: 1-21, 'GGCGSGCGCGSNCGGCGSSCKPVCC', 22-40, 'GSS', 44-45, 'G', 47-48, 'S', 50, 'GSS',
 <W02>
 A:Cross-references: GB:M37759; NID:g200961; PIDN:AAA0106.1; PID:g200962
 A:Note: the sequence reported in this paper has been corrected. See A38660
 C:Superfamily: ultra-high-sulfur keratin

alignment_scores:
 Quality: 427.00 Length: 256
 Ratio: 3.389 Gaps: 24
 Percent Similarity: 49.219 Percent Identity: 37.891

alignment_block:
 US-09-668-119-1/rev x B38346 ..

Align seg 1/1 to: B38346 from: 1 to: 223

```

1016 TCGTCACCTGGGCTGCATCGGCGCTGCACACACATCGAGCTCG 967
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 CysCys.....GlyCysSerGlyGlyCysGlySerCysGlyGly.. 16
966 GACAATTTTCAGTGGTGGTATACACATTTTCACAGGAGAGCTT 917
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 .....CysGlySerSerCysCysGlyProValC 26
916 GCGCGTGCACACCAAGCGTGGTGGTGGTGGTGGTGGTGGTGGT 867
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
26 yscysCysValPro.....ValCysSerCys.....SerSerCysGly 38
866 GGTGGTGTCTGAGCAATGAGAGCTGCTGGGCTGTGTGGCTGCTG 817
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 GlyCysGlyGlyGlyCysGlySerCysGlyGlyCysGlyGly.... 53
816 GTGCTGTGTGTGATGATCTGCTGACGCTGCTGGGAGAGCCYGGAG 767
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 .....CysGlySerCysGlyGlyCysLys..... 61
766 GCGAGAGCTGTGCTGATGATCGTGCATCGTGCATGATGATGCTG 717
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 ..GlyGlyCysGlySerCys...GlyGlyCys...LysGlyGlyCysCys 75
716 TCCAAAGCTGCTGC.....TGCTGCTGC.....TGCTGCTGC 682
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 GlnSerSerCysLys/ProCysCysGlnSerSerCysCysLysPrr 92
681 CTGCTGC.....TGCTGC.....TGCTGC.....TGCTGC 662
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 ocysCysSerSerGlyLysGlySerSerCysCysGlnSerSerCysL 109
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 GGAGCTGCAGCTGCTGATTCGCTGC.....AGCTGCTGT..... 627
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 yspProcysCysGlnSerSerCysLysGlyProcysCysSerSerGly 125
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 .....TGCTGCTGCTGATCTGCTGATTTTGATGATGACAA 589
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 CysGlySerSerCysCysGlnSerSerCysCys..... 136

```

```

588 TTTAATTAGATGCTGCTGC.....TGCTGC.....TGCTGCTGCT 554
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
137 ....LysProcysCysGlnSerSerCysLysProcysCysCysG 152
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
553 GGAGCTGCTGCTGC.....TGCTGACATGCT.....GCTTGAACTGCTGC 513
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 InSerSerCysLysProcysCysSerSerGlySerGlySerSerCys 168
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
512 TGCATGGCACTGCTGAGCGCTGGAACCTGC.....TGCTGCTGC..... 474
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 CysGlnSerSerCys.....CysLysProcysCysCysGlnSe 181
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 ....TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 rserCysCysLysProcysCysGlnSerSerCysCysLysProcysC 198
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 GCTGCTGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 387
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 yscysGlnSerSerCysCysLysProcysCysGlnSerSerCys... 213
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 AGCTGCAGCTGGCTGCTGT 369
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ...CysAlaProValCys 218

```

seq_name: p1r2:T44768
 seq_documentation_block:
 antifreeze glycoprotein AFGP polypeptide precursor [imported] - Boreogadus saida
 C:Species: Boreogadus saida
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T44768
 R:Chen, L.; DeVries, A.L.; Cheng, C.H.C.
 Proc. Natl. Acad. Sci. U.S.A. 94, 3817-3822, 1997
 A:Title: Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid f
 A:Reference number: 222834; MUID:97268653
 A:Accession: T44768
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-507 <CHE>
 A:Cross-references: EMBL:U43200; NID:g2078482; PIDN:AAC60129.1; PID:g2078483
 A:Genetics:
 A:Introns: 1/3

alignment_scores:
 Quality: 399.50 Length: 568
 Ratio: 1.297 Gaps: 20
 Percent Similarity: 54.225 Percent Identity: 29.225

alignment_block:
 US-09-668-119-1 x T44768 ..

Align seg 1/1 to: T44768 from: 1 to: 507

```

184 CTGACGTGGGAGCTGCTGCGGAGCGGCTGGAATGGAGTCCCTCG 233
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LeuThrGlySerCysLeuLeuGlyProSerCysThrSerSerProArgTy 18
234 GGGCCCGGAGACAGTCTCTGGGCGGAGATGATGCTTGTGTCATGGAG 283
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 rgluGlnHisGlnLeu.....ValAla 27
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 AGCCAAATGCTCTCTCAGGAGCGGCTGCTGAGACTCGGGGATGGCC 333
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
27 rglProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 40
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 CCTCAGCAGTGGCTGCTGTCTAGGCACTGCACAGCCACAGCTCA 383
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 .....AlaThrProAlaThr..AlaIa 47
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 GCTCCAGAGTGGCGCTGCACAGCA.....GGAGCAACAGC 421
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

48 ThrPrAlaIeThrAlaAlaIeThrAlaAlaIeThrgluAlaIeThrAlaIeThrAl 64
422 AGCAgTTCACGAGCAGCAGCAGCGGGCGCTACGACGACGACGACGA. 470
|||:::|||||:::||||| |||:::|||||:::|||||
64 aaIaThrProAlaIeThrAlaIeThrAlaIeThrProAlaIeThrAlaIeThr 81
|||:::|||||:::||||| |||:::|||||:::|||||
471GCAGCAAAcGACACACTCCAGCGGTGCGAGAGTGCCATGcAGA 515
|||||:::|||||:::|||||::: |||
81 hrThAlaIeAlaIeThrAlaIeAlaIeThrAlaIeAlaIeThrProAlaIArg 97
|||:::|||||:::|||||::: |||
516 GCAGTTCACAAGCAGTAGTGCAGCAGCAGCAGCACAGCTCCGACGAGCAGC 565
|||:::|||||:::|||||::: |||
98 AlAaIaIaIaIaIaIaIeThrProAlaIeThrAlaIaIaIeThrProAlaIeAl 114
|||:::|||||:::|||||::: |||
566 AGCAcGACAGCAGCATCTAAATTAAATGCATCAAAA..... 602
|||:::|||||:::|||||::: |||
114 aThrAlaIaIeThrAlaIaIaIeThrAlaIaIeThrAlaGlutThrProAlaIArg 131
|||:::|||||:::|||||::: |||
603TCACAACAAGTAACGAGCAGCAGCAACGACGAGCT 635
|||:::|||||:::|||||::: |||
131 IaAlaIeThrProAlaIeThrAlaIaIaIeThrProAlaIeThrAlaIaIeThrProAla 147
|||:::|||||:::|||||::: |||
636 GCAGCAATAGCAGCAGCTGCAGCTCCAACAACAAGCAGCAGCAGCAGCAGC 685
|||:::|||||:::|||||::: |||
148 ThraIaIaIeThrAlaIaIaIeThrAlaIaIeThrSerAlaIeThrAlaIaIeThrAl 164
|||:::|||||:::|||||::: |||
686 AGCAcGACAGCAGCAGCAGCAGCGGTTGGAGGCCCGACCAATTGAG 735
|||:::|||||:::|||||::: |||
164 aaIaIaIaIaIaIeThrProAlaIeThrAlaIaIaIeThrProAlaIeThrProAlaI 181
|||:::|||||:::|||||::: |||
736 CAGCCACGAGTGCACCA.....GCCACGCGCTCGGGCCTC 770
|||:::|||||:::|||||::: |||
181 hrAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIeThrProAlaIeThrAlaIeThrAla 197
|||:::|||||:::|||||::: |||
771 CCAGGCTGTGCCCCAGCAGCTGCAGCAGCAGATGCATCAGCAGCAGCAGCAGC 820
|||:::|||||:::|||||::: |||
198 AlaIeThrAlaIaIaIaIeThrAlaIaIaIeThrAlaIaIaIeThrAlaIaIeTh 214
|||:::|||||:::|||||::: |||
821 AGCCGCGACACAGCGCCCAGCAGCGCTCCAGTTGCTCAGAACCAACATCA 870
|||:::|||||:::|||||::: |||
214 rAlaIaIeThrProAlaIaIaIaIaIaIaIaIaIaIaIeThrProAlaIeThrAlap 231
|||:::|||||:::|||||::: |||
871 CAACGCGCGCACAGCTGCAGACCCAGCGCTTGGGTGCAGCGGGCGAACG 920
|||:::|||||:::|||||::: |||
231 roThrProAlaIeThrAlaIaIaIeThrProAlaIeThrAlaIaIaIeThrAlaIeThr 247
|||:::|||||:::|||||::: |||
921 TCTCCCTGGACAAATGTGTATACCACAACCAACACTGAATTTGCCGAG 970
|||:::|||||:::|||||::: |||
248 AlaPro.....ThraIaIaIeThrProAl 255
|||:::|||||:::|||||::: |||
971 CTCGCAATGGGTGCAGAGCCCGCAGT..... 998
|||:::|||||:::|||||::: |||
255 aaIrgAlaIaIaIaIaIaIeThrProAlaIeThrAlaIaIeThrAlaIeThrAla 272
|||:::|||||:::|||||::: |||
999GCAGCCCAAGGTGCAGCAGCAGCAGCAGCAGC 1028
|||:::|||||:::|||||::: |||
272 IaAlaIeThrProAlaIeThrProAlaIeThrProAlaIeThrAlaIaIeThrAspAla 288
|||:::|||||:::|||||::: |||
1029 ACtAcAGAGAGCTGAGGCTGCCA.....GA 1054
|||:::|||||:::|||||::: |||
289 ThraIaIaIeThrAlaIaIaIeThrProAlaIaIaIaIaIeThrProAlaIeThrPr 305
|||:::|||||:::|||||::: |||
1055 TGGTGGCTCCCGAGTCCAGGTCAAGTCACGAGAGAGCTCCCGATGCTGCC 1104
|||:::|||||:::|||||::: |||
305 oAlaIeThrAlaIaIeThrProAlaIeThrProAlaIeThrAlaIaIeThrAlaIeThr 322
|||:::|||||:::|||||::: |||
1105 TCAGCGTCACCGGCGCAGCTGCAGACCCCGCAGTGCATGCC..... 1148
|||:::|||||:::|||||::: |||
322 hrAlaIaIaIaIaIaIaIeThrAlaIaIaIeThrProAlaIaIaIaIaIaIaIaIaIaIa 338
|||:::|||||:::|||||::: |||
1149CCGTCCCGCCAGCGCTCCCGCAGCGCGGCGGCGACGCGCTGAC 1192
|||:::|||||:::|||||::: |||
339 AlaIeThrProAlaIeThrAlaIaIaIeThrProAlaIeThrAlaIaIaIaIaIaIaIaIaIaTh 355

seq_name: p1r2:A56158

seq_documentation_block:
eye development protein canoe - fruit fly (Drosophila melanogaster)
C|Species: Drosophila melanogaster
C|Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 21-Jul-2000
C|Accession: A56158
R|Miyamoto, H.; Nihomatsu, I.; Kondo, S.; Ueda, R.; Togashi, S.; Hirata, K.; Ikegami
Genes Dev. 9, 612-625, 1995
A|Title: canoe encodes a novel protein containing a GLGF/DHR motif and functions with
A|Reference number: A56158; MUID:95212910
A|Accession: A56158
A|Status: preliminary; not compared with conceptual translation
A|Molecule type: mRNA
A|Residues: 1-1893 <MUY>
C|Cross-references: GB:D49534; NID:g705386; PIDN:BAA08478.1; PID:g705387
F:848-924/Domain: GLGF domain homology <GLG2>

alignment_scores:

	Quality:	Ratio:	Length:
Percent Similarity:	399.00	1.308	664
Percent Identity:	45.934	27.410	30

alignment_block:
US-09-668-119-1 x A56158 ..

Align seg 1/1 to: A56158 from: 1 to: 1893

```

19 GCACACAGTAATCCACAGAGATATGAGAGCCATGTTTCTGTGAAGGC 68
   |||||
1010 AAlaHisLeuProAsnSerLysSerValProAla.....LeuHisNI 1023
69 CAAGACCCGGGACGAATACCTTTCTCTGAGCCAGGCTCATTTATCAT 118
   |||
1023 HisThrGlySerGlyThrIleSerLeuAla..... 1033
119 TTTCGAGACATTCATTAAGAAATCTCAAGCTCCGTCATGATGATCCATG 168
   |||||
1034 .....AsnSerLysSerAlaGlySerThr..... 1040
169 AATGACATCCAGAGCCTGACTGGCGGACCTGCT.....GGGGAGCCGC 212
   |||||
1041 HisSerLeuHisAsnAsnThrSerIleMetGlyGlyIleGlyGlyAlaAl 1057
213 TGGAAATGCGCATG.....CCTCCTCGGGGCCGGGACAGTCTCTGGGCG 256
   |||
1057 AlaIleGlyIleMetLeuIleGlyIleProAsnGlySerGlnAsnAlaAsnG 1074
257 GGATGGGTAGC..... 270
1074 IAsnGlyIleAsnGlyIleAsnGlyIlePheTyrGlnAsnLeuSerValTyr 1090
271 GGTGCCATGGGACAGACCAATGTCTCTCTCAGGAGCGCCGCTCCTGGGAC 320
   |||||
1091 ArgAlaGlnAsnGlnSerGlnProIleLeuAsnGlyIleArgProProIleAl 1107
321 CTGGGGGATGGCCCTACAGCATGGCTGTGCTGCTCTACGGCAGCATGCCAC 370
   |||
1107 AlaHisAlaIleMetAsnAlaTyrAsnGlySerSerProLeuAlaProG 1124
371 AGACCCACAGCTGAGCTACAGAGTGGGGCTGCGACGACAGCAGCAGCAG 420
   |||
1124 GlnGlnIleProGlnIleGlnGlnIleSerProTyrGlnGlnGlnGlnGln 1140
421 CAGCAGTCCAGCAG..... 435
1141 GlnGlnIleHisMetGlnAlaAsnAlaAsnLeuProProThrArgProValSe 1157
436 .....CAGTAGCAGCGCGGCTACAGCAGCAGCAGCAGCAGCAGC 472
1157 AlaTyrTyrHisSerIleGlnIleSerAlaGlnGlnGlnLeuGlnGlnGlnG 1174
473 AGCAACAGCAGCAGCTTCAGGCTCAGCAGAGTGGCATGACAGCAGCAGTTC 522
   |||||
1174 GlnGlnGlnGlnGln.....GlnHisSerIleGlnGlnGlnGlnGlnPhe 1187
523 .....CAAGCAGTAGTGCAGCAGCAGCAGCAGCAGCT 551
1188 AlaLeuSerSerGlyIleLeuAsnGlyGlnGlnGlnGlnGlnGlnGlnGln 1204
552 CCAGCAGCAG.....CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 574
   |||||
1204 sGlnHisGlnLeuThrLeuAsnAsnArgThrLysSerGlnGlnAsnIle 1221
575 AGCATCTTAATTAATTT..... 591
1221 LHisThrLeuArgMetGlnGlnIleMetAlaProSerMetProAsnIle 1237
592 .....CATCATCAAAATCAGCAGCAGCAG.....ATACA 617
1238 SerAsnMetTyrHisIleGlnGlnGlnHisGlnGlnGlnGlnGlnGlnGln 1254
618 CCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 642
   |||||
1254 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1271

```

```

642 ..... 642
1271 etGlnAsnValAsnAspPheAlaGlyIleTyrGlnAsnGlySerLeuGln 1287
643 .....ATAGCACA 650
1288 TyrArgArgSerGlnLeuHisAspProSerThrLeuTyrGlnIleGlnIle 1304
651 GCTGCAGCTCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 700
   |||||
1304 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1321
701 AGCAGCAGCT.....TTGAGGCCAGCAGCACA..... 729
1321 LHisAlaSerProAsnPheIleAlaLeuProProIleProLysProLeuGlySer 1337
730 ATTTCAGACGCCAGATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT.. 777
   |||||
1338 LeuGlnSerPro.....AsnLysProAsnValProProSerThrAlaIle 1352
778 .CTGCCACAGCAGCTCAGCAGATGCATCAGCAGCAG.....C 814
   |||||
1352 oLysProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1369
815 ACCACAGCGCCGACCCAGCAGCC..... 837
1369 LAspLysProProLeuProProIleProThrAlaThrHisProLeuPheLysAla 1385
838 .....CAGCAG.....CC 845
1386 ThrGlnGlnIleAlaIleProGlyMetAsnTyrValAlaSerThrLeuAspIle 1402
846 TGCA.....GTGCTCAGAACCAACCAACCAACCAACCAACCAACCAACCAAC 883
   |||||
1402 oProLysGlySerTyrValAlaIleSerAsnGlnGlnGlnGlnGlnGlnGln 1419
884 AGTCGACAGCCAGCCT..... 900
1419 IserGlySerAsnProTyrGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1435
901 .....TTGCTGTCAAGGC 914
1436 MetArgArgGlnHisIleLeuArgIleThrArgGlnGlnGlnIleSerGln 1452
915 GCAAGCTCTCCCTGGACAAATGTTGATACCCACCAACCACTGAAA.... 960
   |||||
1452 uSerGlnIleValIleSerArgSerProMetGlnGlnGlnGlnGlnGlnGln 1469
961 .....TTTGTCCAGCTCGATGGTGTGCAGCAGCC 993
1469 euIleLeuGlnArgAspPheGlnArgAlaGlnGlnGlnGlnGlnGlnGln 1485
994 CCAGTGCAGCCCGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1043
   |||||
1486 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1500
1044 GGCTGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1093
   |||||
1500 nGlnLeuPheArgLeuAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1517
1094 CCAATGCTGCTCTCCGCTCAGCAGCGGGCCAGCAGTGCAGACCCCGCAGTCC 1143
   |||||
1517 roIleThrSer.....TyrArgGlnHisGlnIleLys 1527
1144 ATGCCCCCTCCCGCCAGCGCTCCCGCCAGCGCGCCGACCCAGCAGCTCACA 1193
   |||||
1528 LeuAlaGlnMetPro.....As 1533
1194 GCCCAACTCCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1243
   |||||
1533 pSerAsnSerLeuValAspSerValPro...ProGlnPro..... 1545
1244 TGCCAGCCCTCAGCAGCAGCC.....TCCAGAGCCAGCTG 1281

```



```

1393 CAGTACTGCGACAGCTGAAGCAGCTGGCAATCATGACATGAGCCCTGGG 1442
      ::::: ||| |||||:::
539 ..... SerSerIarLysAlaIalaserSergLyAsnGLUserIetSe 552
1443 CGGCATGATTCACAAC.. ATCGACAAGAAGCA..... 1473
      ||:::|||||:::|||||:::|||||
552 rASpHisIleSerArgIeIetSerGIAsngInValIleueGlnGLYA 569
1474 .....CACAGAAAMAG..... 1485
      |||||
569 sPProVALIleArgLYSLysArgProTYrHisArgGlnIleGLYAlaGln 585
1485 ..... 1485
586 SerSerValASPHisApsSerAnsSerGLYSerThrArgThrserPr 602
1485 ..... 1485
602 oGLYProLYsAspSerArgMetLeuGlnAlaIalaserArgSernIerL 619
1486 .... GACCTGAGTAATATGAAGACCTCTGGACATTGTGACA..... 1524
      :::||||| ||:::|||||
619 euRHeLIueuSerGLYSerLYSHisPhemeGLYserIethrSergLY 635
1525 ..... GACCCTGCCAAGCGG... 1539
636 GLNProLeuLeuArgProIleGlnAlaHisAsnSProAsnIyrThrPr 652
1540 .....TGTTCCCTG..... 1548
652 ocLUcysIleTYrCYsLYsLeuThrPheProAsngInValagLYeuGlnA 669
1549 ..... AGACCTTGCAAAAAGTGTGAGATC 1572
669 lAnISlcyValALCySGLYLYSLySGluLeuGLuIucyslaGlnIle 685
1573 GGCTTGGAATAACTCAAGAATGACATGCGGTCGCCACTGCCACCGCG.. 1620
      || | ||
686 AlAGlnGLU.....GLYAsnProHisseral 694
1621 .... CCCCGTGGCCACTGACCAACAGCAGTACCTTGCCAGCGCGCTGT 1666
      ||| |||:::||||| |||||
694 alEuLYsAtgArgHisThrHisGlnAspAlathrIleuAlaMeHisSer. 710
1667 GGATGGCGTGGCCACATCGCGCTACCTGTCTCAACATATGCC 1713
711 .....ProLeuAlaAlaHisThrProSerAsnMetPro 721

seq_name: pIR2:TI3998
seq_documentation_block:
gene mastermind protein - fruit fly (Drosophila virilis)
C:Species: Drosophila virilis
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: J13398
J:Newfeld, S.J.; Tachida, H.; Yedvobnick, B.
J.Mol. Evol. 38, 637-641 1994
A:Title: Drive-selection equilibrium: homopolymer evolution in the Drosophila gene mastermind
A:Reference number: Z17850; MUIDD:94365848
A:Accession: J13398
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-1655 <NDW>
A:Cross-references: EMBL:H92914; NID:g15f7833; PID:g15f7834; PIDN:AAC37201.1
A:Genetics:
A:Cross-references: FlyBase:Fbgm0013119
```

[illegible]

```
alignment_block:
US-09-668-119-1 x T13998 ..
Align seg 1/1 to: T13998 from: 1 to: 1655
```

[illegible]


```

640 .....CGA 642
887 rglyProGlnSerAsnProAsnAlaValProGlyGlyAsnAlaAlaAsn 903
643 ATAGCAGCACTGCTGAGTCCACAGACAGCAAGCAGCAGCAGCAGCAGCA 692
904 AlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 920
693 CGACGACGACGAGCTTTGGAGGC..... 720
920 ngInGlnGlnGlnGlnGlnGlnGlnAlaThrThrThrLeuGlnMetL 937
720 ..... 720
937 ysGlnThrGlnGlnLeuHisLeSerGlnGlnGlyGlySerHisGly 953
720 ..... 720
954 IleGlnValSerAlaGlyGlnHisLeuHisLeuSerSerAspMetLysSe 970
721 .....CAGCCAC 727
970 rAsnValSerValAlaAlaGlnGlnGlyValPhePheSerGlnGlnAla 987
728 CAATTGACGACGACGACGACGACGACGACGCTCCGCTCCAGGCT 777
987 laAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1003
778 CTGCGCCGACGCTGACGACGATGACACACACACAC..... 816
1004 AsnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1020
817 .....CACCAGCCGACACAGCCGACGACGCT..... 846
1020 nglyGlyGlyProAsnGlyProGlnGlnGlnGlnGlnGlnGlnGlnGln 1037
847 .....CCAGTGGCTGACGAACCA 867
1037 snAsnSerAsnValProSerAspGlyPheSerLeuSerGlnSerGlnSer 1053
868 TCACAACTCCGCGCAGCAGCAGCAGCAGCAGCTTGTGTGACAGCGCA 917
1054 MetAsnPheThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1070
918 AGCTCTCCCTGGACAAATGTTGATATACCAACACACACTGAATTTGCTCC 967
1070 alaAlaAlaAlaGlnGlnGlnGlnGlnAla..... 1079
968 GAGCTCCGATGGTGTGACAGCAGCCCACTGACGCGCCAGTGCACAG 1017
1080 .AlaAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1095
1018 CAGCAGACAGCAGTACAGCAGCAGTCCAGTCCAGATGGTGTCCCGG 1067
1096 ArgGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1110
1068 AGTCAGGTCAGCAGACAGCAGCTCCCACTGCTCCCTCCGCGTACCGG 1117
1111 .....AlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1125
1118 GC.....CAGCAGGTCAGCAGCCCGCAGTCCGATGATG 1146
1125 lyGlyAsnValProLeuMetGlnGlnGlnGlnGlnGlnGlnGlnGln 1138
1147 CCCCCCTCCCGCAGCCGCTCCCGCAGCCGCGCAGCCAGCAGCAGCAGC 1196
1139 .....GlyGlyValProValGlyAlaGln 1146
1197 CAATCCCAAGTCAGCTGCTGCGCCGCGCCATCTCCAGTGTTCCTGCG 1246
1146 yserGlyAsnAlaSerValGly..... 1153

```

```

1247 CCAGCCCCCTACCCGACGCTCCCGACAGCCCACTGACGCGCGGACCCCA 1296
1154 .....ValProValSerAlaGlyGlyPro 1161
1297 CAGAACTTCACT..... 1308
1162 AsnAsnGlyAlaMetAsnGlnLeuGlyGlyProMetGlyGlyMetProGln 1178
1309 .GTCCCTCCCTGACCTTTAAACACACCTGTGACCCGACGCTGTGCA 1357
1178 yMetGlnMetGlyGlyProGlyGlyValProLeuAsnProMetGlnMetL 1195
1358 TGAGCCGACGTGGCTCCGACGAGCTGA..... 1385
1195 snProAsnGlyGlyAlaProAsnAlaGlnMetMetGlyGlyAsnGly 1211
1386 GGAGCAGCAGTACCTGACAAAGCTGACAGCCTGTC.....GA 1423
1212 GlyGlyProValProAlaAlaSerGlnAlaLysPheLeuGlnGlnGln 1228
1424 ACTACATGAGCCCTGCGCGCATGATCAACAGATGACACAGACGAA 1473
1228 nLeMetArgAlaGlnAlaMetGlnHisGln..... 1239
1474 GACAGAAAAAGCAGCTGATGATGAAGACCTTGTGACATTTCTGAC 1523
1239 ..... 1239
1524 AGACCCCTCGAAGCGGTGTCCTCGAAGACCTTGCAAAAGTGTGAGATG 1573
1239 ..... 1239
1574 CCTGTGAGAACTCAAGATGATGCGGTGCGCATCCCGCCGCGCCG 1623
1240 .....GlnValGlnGlnHisMetAlaGlyAlaArgProProProG 1254
1624 CGGTGCCACCGACCAACAGCAGTACTTA 1652
1254 lntyrAsnAlaThrLysAlaGlnLeuMet 1263

seq_name: p1r2:S25365
seq_documentation_block:
CYC8 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: glucose repression mediator; protein YBR0908; protein YBR112c; SSN
C:Species: Saccharomyces cerevisiae
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000
C:Accession: S25365; S48277; S45980; S25404; S25405; A30906; S44692
R:Mannhaupt, G.; Stucka, R.; Ehmler, S.; Vetter, I.; Feldmann, H.
Yeast 8, 397-408, 1992
A:Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excisio
A:Reference number: S25364; MUID:92327848
A:Accession: S25365
A:Molecule type: DNA
A:Residues: 1-966 <MAX>
A:Cross-References: EMBL:X66247; NID:93548; PIDN:CAA46973.1; PID:93550
R:Mannhaupt, G.; Stucka, R.; Ehmler, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357
A:Accession: S48277
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-966 <MAX>
A:Cross-References: EMBL:X78993; NID:9476045; PIDN:CAA55615.1; PID:9476068
R:Feldmann, H.; Mannhaupt, G.; Schwarze, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45980
A:Molecule type: DNA
A:Residues: 1-966 <FB2>
A:Cross-References: EMBL:Z35981; NID:9536449; PIDN:CAA85069.1; PID:9536450; MIPS:YBR1

```

R:Schultz, J.; Carlson, N.
Mol. Cell. Biol. 7, 3637-3645, 1987
A:Title: Molecular analysis of SSN6, a gene functionally related to the SNF1 protein kin
A:Reference number: S25404; M0ID:88065502
A:Accession: S25404
A:Molecule type: DNA
A:Residues: 1-546, 'K', 546-966 <SCH>
A:Cross-references: EMBL:M17826; NID:g172725; PIDN:AAA35103.1; PID:g172726
R:Trumbly, R.J.
Gene 73, 97-111, 1988
A:Title: Cloning and characterization of the CYC8 gene mediating glucose repression in Y
A:Reference number: S25405; M0ID:89211964
A:Accession: S25405
A:Molecule type: DNA
A:Residues: 1-546, 'K', 546-966 <TRU>
A:Cross-references: EMBL:M23440; NID:g171349; PIDN:AAA34545.1; PID:g171350
A:Gene: S6D:CYC8; SSN6; CRT8
A:Cross-references: SGD:S0000316; MIPS:YBR112c
A:Map position: 2R
C:Function:
A:Description: required for complete derepression of ICU1; required for repression of SU
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
C:Keywords: nucleus; transcription regulation
F:224-257/Domain: tetratricopeptide repeat homology <TT1>
F:262-295/Domain: tetratricopeptide repeat homology <TT2>
F:296-329/Domain: tetratricopeptide repeat homology <TT3>
F:330-363/Domain: tetratricopeptide repeat homology <TT4>
F:365-398/Domain: tetratricopeptide repeat homology <TT5>

alignment_scores:
Quality: 386.50 Length: 486
Ratio: 1.487 Gaps: 16
Percent Similarity: 53.498 Percent Identity: 29.630

alignment_block:
US-09-668-119-1 x S25365 ..

Align seg 1/1 to: S25365 from: 1 to: 966

```

280 GGACAGCCCATGTCT.....CTCTGAGGAGCCCGCTCTGAGCAGTC 323
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 GLYAsnProleuAsnThrArgIleSerAlaGlnSerAlaAsnAlaThrAl 463
324 GGGGATGGCCCTCACAGCATGGCTGCTGCTCTAGC..... 360
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
463 AsnMetValGlnGlnGlnHisProAlaGlnThrProIleAsnSer 480
361 .....GCACProCAGAGCAGCCGAGCTGAGCTC 387
    |||:|||||:|||||:|||||:|||||:|||||:
480 eAlaThrMetLysSerAsnGlyAlaSerProGlnLeuGlnAlaGlnAla 496
388 CAGCAGAGTGGCGCTGCAGCAGCAGCAGCAGCAGCTTCCAGCAGCA 437
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
497 GlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGln 513
438 GAGCAGCAGCGCGCTTACAGCAGCAGCAGCAGCAGCAGCAGCAGT 487
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
513 nAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 530
488 TCCAGGCTTCA.....CAGAGTGCATG 510
    |||:|||||:|||||:|||||:|||||:|||||:
530 lAlaAlaGlnAlaGlnAlaAlaHisAlaGlnAlaGlnAlaGlnAla 546
511 CAGCAGCAGATTCCAGCAGTAGTAGCAGCAGCAGCAGCAGCTCCAGCACA 560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
547 GlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGln 563
561 GAGCAGCAGCAGCAGCAGCTAATTAATTCAGCAGCAAAATTCAGCAGC 610
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
563 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 580

```

```

611 AGATACAGCAGCAGCAGCAGCAGCTTGCAGCGAATAGCAGCAGCTGAGCTC 660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
580 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 596
661 CAACAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 710
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
597 GlnGlnLysGlyValSerValGlnMetLeuAsnProGlnGlnGlnGln 613
711 TTGGAGGCGCCAGCCCATTCAGCAGCAGCAGCAGCAGCAGCAGCAGCA... 756
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
613 oTyrIleThrGlnProThrValIleGlnAlaHisGlnLeuGlnProPhe 630
757 .....CAGCCTCCGCGCTCCAGCGCTGCGCC..... 783
    :|||:|||||:|||||:|||||:|||||:|||||:
630 eThrGlnAlaMetGlnHisProGlnSerSerGlnLeuProGlnGln 646
784 CAGCAGCTGAGCAGCAGATGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 833
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
647 GlnGlnLeuGlnSerValGlnHisProGlnGlnGlnGlnGlnGlnProG 663
834 GCCCAGCAGCAGCT...CCAGTTGCTCAGACCAACCATCACA.....C 874
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
663 nAlaGlnAlaProGlnProleuIleGlnHisAsnValGlnGlnGlnVal 680
875 TCCCGCCACAGTGCAGACCCAGCCTTGGTG...TCACAGCGCCAGCT 921
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
680 euProGlnLysArgTyrMetGlnGlyAlaIleHisThrLeuValAspAla 696
922 CTCCTGCAACAATGTGTATACCCACACACACCTGAATTTGTCGAGC 971
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
697 AlAlaValSerSerThrHisThrGlnAsnAsnThrLysSerProArgG 713
972 TCCG.....ATGGTGTGCAGACCCCA.....G 997
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
713 nProThrHisAlaIleProThrGlnAlaProAlaThrGlyLeuThrAsn 730
998 TGACAGCCCGCAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 1047
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
730 lAluProGlnValLysLysGlnLysLeuAsnSerProAsnSerHisIle 746
1048 GCCCAGAGTGTGCTCCCGAGTCCAGTCCAGCAGCAGCAGCAGCAGCTCCAT 1097
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
747 AsnLysLeuValAsnThrAlaThrSerIleGlnLysAsnAlaLysSerG 763
1098 GCTGTCTCGCCGCTCACCGGGCCAGCAGGTGCAGACCCCGCAGTCGATGC 1147
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
763 uValSerAsnGlnSerProAlaValValGlnSerAsnThrAsnAsn... 778
1148 CCCCTCCCGCCAGCAGCTCCCGCAGCCGCGCCAGCCAGCTCAGCAGCC 1197
    :|||:|||||:|||||:|||||:|||||:|||||:
779 .....ThrSerGlnGlu 782
1198 AACTCACAGTCACTGTGCGCTGCCCAATTCCTCCAGTCTCCGCGC 1247
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
783 GluLysProValLysAlaAsnSerIlePro.....SerValIleG 796
1248 CAGCCCTCAACCGCAGCTCCAGCAGCCAGTGCAGGCGCGGAGCCAC 1297
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
796 yAlaGlnGlnProProGlnLysAlaSerProAlaGlnGlnGlnAlaThrLys 813
1298 AGAAGTTCAGTGTCCCTCAGCTGAGCTTAACACACACCTGTGAACCC 1347
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
813 lAlaLysValSerProSerThrLysProLeuAsnThrGluProGlnLys 829
1348 AGCTCTGTCAATG.....AGCCAGCTGCTCCAGCCAGCAGCTGA 1385
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
830 SerSerValGlnProThrValSerSerGlnSerSerThrThrLysAlaAs 846
1386 GAGCAGCAGTACCTGAGCAGAGCTGAAGCAGCTGTGCAAGTACATGAGAC 1435
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
846 nAspGlnSerThrAlaGlnThrIle...GlnLeuSerThrAlaThrValP 862
1436 CCTT.....GGCCGCGATGATCAACAAGATGACAGCAAGCAAGACAGCAG 1478

```


Fri Mar 1 09:11:59 2002

us-09-668-119-1.rpt

Page 14

[illegible]

seq_documentation_block:
 G-box-binding factor - slime mold (*Dictyostelium discoideum*)
 C:Species: *Dictyostelium discoideum*
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
 C:Accession: A53185
 R:Schmützler, G.R.; Fischer, W.H.; Firtel, R.A.
 Genes Dev. 8, 502-514, 1994
 A:Title: Cloning and characterization of the G-box binding factor, an essential component of the *Dictyostelium* cell cycle
 A:Reference number: A53185; MUID:94170994
 A:Accession: A53185
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-708 <SOCH>
 A:Cross-references: CB:123075; NID:g456561; PIDN:AAA21021.1; PID:g456562
 C:Superfamily: G-box binding factor
 C:Keywords: DNA binding; transcription factor; zinc finger

alignment_scores:	Quality: 383.00	Length: 6827
	Ratio: 1.212	Gaps: 23
Percent Similarity:	43.185	Percent Identity: 24.390
alignment_block:		
US-09-668-119-1 x A53185		..

Align seg 1/1 to: A53185 from: 1 to: 708

```

130 CATTAACAAGAAATGCTCAAGTTCCGTACCGTAACTCATGGATTCGCACATCCA 179
131 ..... : : : : : | | | | | : : : : : ||
6 HistHingInglYanSerSerSerSerSerSerSerSerSerProseogl 22
180 GAGCCTCACACGCGGGAGACCTGTCCGGGAGGCCCTTGGAATTGGCATGGCTC 223
181 ..|||..: : : : : | | | | | : : : : : ||
22 nPrHleelyYelYerAspleuSerAsnIleSerAlaLeuProLeuProL 39
230 CTCGGGGCC..... 237
39 euProSerIlePheThrThrAlaGlnAsnGlnMetAsnProProIleLeu 55
238 ....'CCGGGACAGACT...CTGGCGCGGATGGGTACGCTTTGGTGGCAT 278
239 ||||| ||||| ||||| ||||| : : : : : ||
56 PheProThrSerSerLeuLeuGlyGlySerSerAsnThrProserPh 72
279 GGAGCAGCCAAATGCT....CTCCAGAGGAGCGCGCTCTGGAGACT 322
280 ||| ||| : : : : : | | | | | : : : : : ||
72 eleLeuProProSerSerIleMetSerSerAsnValPheProSerHisA 89
323 CGGGGATGGCCCCCTCACAGCATGGCTGTGCTCT..... 357
89 spgLygInlYrProAsmMetProAsmMeValAspGIrPlutyrGlnIleHis 1057
358 ...ACGGCAACTCCACAGACCAGCTCCAGCTCCAGAGGTGGGCGCTGCA 404
106 ProAsnGlnAsnProHisTyrrAsnYrGlnItyrGlnIleuMetPrMetGl 1222
405 GCACAGCAGCAGAACACAGCAGACTCTCCAGCAGCAGCAGCGCGCTAC 454
406 ||||| ||||| ||||| ||||| : : : : : ||
122 ngdInAlaIgLnAsngLnProProGlnLnAsngLn.....G 136
455 AGCAGCAGCAGCAGCAGCAGCAAAGCAGCAAGCTCCAGCTCCAGCAGAGT 504
136 InGlnGlnInHisIgLnGlnGlnGlnGlnGlnProGlnInHisIgLnGln 152
505 GCCATGCAGCAGCAGTCCAGCAGACTAGTCCAGCAGCAGCAGCAGCTTCA 554
153 MetGlnInGlnInHisIgLnGlnGlnMetGlnGlnGlnGlnGlnInHisI 1599
555 GCACAGCAGCAGCAGCAGCAGCAGCATCTAATTAAATTGCATCATCAAAATC 604
169 sgInGlnMetGlnGlnGlnGlnInHisIgLnGlnMetGlnInHisIgLnL 186
605 AGCAGCAGATACAG.....CAGCAGCAGCAGCAGCTCCAGCAGATA 645
186 euGlnGlnInHisIgLnInHisIgLnGlnGlnGlnGlnGlnGlnGlnInHis 202
646 GCACAGCTGAGCTCCAAACAGCAGACAGCAGCAGCAGCAGCAGCAGCA 695
203 GlnGlnGlnInHisIgLnGlnGlnGlnGlnGlnGlnGlnGlnGlnInHisIg 219
696 GCACAGCAGCAGAGCTTTGGAGGCCAGCAGCAGCAATTCCAGCAGGCCAGA 745
219 ngGlnGlnGlnInHisIgLnInHisSerGlnProGlnGlnGlnInHisIgLnIsA 236
746 TGCCAGCAGCCACAGCTCCGCCCTCCAGGCTCCGCCAGCAGCTCCAG 795
236 sngInGlnGlnGln.....HistGlnInHisAgLn 245
796 CAGATGCTACACAGCAGCAGCAGCAGCCGCGCAGCAGCAGCCAGCAGCC 845
246 GlnGlnInHisIgLnGlnGlnGlnAsngLnIleGlnMetValProGlnInPr 262
846 T..... 846
262 ogInSerLeuSerAsnSerGlyAsnAsnAsnAsnAsnAsnAsnAsnA 279
846 ..... 846
779 snAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnSerInHisGln 295

```



```
232 .....CGGGGCGCCG 240
622 roHISglMetGlnGlnGlyLysProMetSerMetAsnArgProPro 638
241 GGACAGTCTCTGGGCGGATGGTAGCTTGTGTCATGGACAGCCCAAT 290
639 SerGlnMetHisGlnGlnGlyLysProGlnLeuMetAsnArgProse 655
291 GTCTCTCTCAGGCGACCGGCTCTGGAGCTGGGATG..... 330
655 rSerGlnMetGlyLysGlnProGlnGly.....GlyMetMetSerArgp 670
331 .....GCCCCCTCAGCAGCATGGCTGTC 351
670 roProSerGlnProHisGlyLysGlnGlnGlyProSerSerMet..... 684
352 GTGTCTAGGCACTCCAGACCCAGCTGACGTCAGCAGGTGGCG... 399
685 .....AsnThrProGlnGlnGlnMetArgMetGlnGlnGlnGln 699
400 .....CTGCAGCAGCAGCAGC 415
699 nasAsnGlnAsnMetAsnGlnGlySerProHisGlnGlnGlnProLeuG 716
416 AACGACGAGCTGCCAGCAG.....CAG 438
716 LHisGlnGlnMetGlnGlnGlnAsnAlaProAsnMetAsnMetProGln 732
439 CAGCAGGCGGCGCTACAGCAGCAGCAG..... 465
733 GlnGlnLysArgGlyLysGlnGlnGlnProGlnMetLeuGlnGlnGln 749
466 .....CAGCAGCAGCAGCAGCAGCAGCAGCTTCAGC 493
749 rAsnGlyProProProProHisGlnGlnGlnGlnGlnProValArgV 766
494 CTGACGAGAGTCCATZCAGCAGCAGCTCCAGCAGTAGTCAGCAGCAG 543
766 alAlaGlnValAlaGlnGlnProGlnAsnMetGlnLysProAlaArgV 782
544 CAGCAGCTCCAGCAGCAGCAG..... 564
783 MetThrProGlnGlnSerGlnSerLysGlnGlnGlnSerMetGlnAsnSerPr 799
565 .....CAGCAGCAGCAGC 577
799 ogLHisProGlnMetArgMetAsnLeuAlaProProGlnGlnGlnGln 816
578 ATCTAATTAAATGCAICATCAAAATCAGCAACAGTACAGCAGCAGCAA 627
816 L.....HisGlnAsnGlnProProMetGlnGlnGlnAsnPro 828
628 CAGCAGCTCAGCAGATACAGCAGCTGACGTCACAAACAGCAA..... 672
829 GlnSerAspArgGlnMetProGlnMetGlnGlnProGlyArgGlnLysHis 845
673 .....CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTT 712
845 sProAsnLysArgGlnProGlnHisGlnMetGlnGlnAsnArgSerMet 862
713 TGGAGGCCAGCA.....CCAATTCAGCAGCCAGCCAG 744
862 etGlnGlnLysProHisAsnProGlnAsnLysProAspGlnMetMetAsn 878
745 ATGACAGCAGCAGCAGCTCCGCGCC.....TCCAGAGCTCTGCCCCAGCA 788
879 MetGlnGlnMetGlnGlnAspProMetSerGlnGlnProLeuGlnGlnLys 895
1789 GCTGACAGAGATGATCAGACAGCAGCAGCAGCCAGCCAGCA..... 831
895 rProGlnProAsnLeuHisLysGlnHisGlnHisGlnThrProProMetGln 912
832 .....CAGCCG..... 837
912 etGlnMetMetGlnProValValThrAspAsnPheGlnProAspAsnAsp 928
837 ..... 837
929 GlnThrLysArgLysAsnGlyLysHisAlaLysLysGlnGlnGlnGln 945
838 .....CAGCAGCTCCAGCTG 853
945 rLysLeuValProGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 962
854 CTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 876
962 GlnGlnGlnGlnProMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 978
877 .....CCGCCAGCTCCGAG.....ACCA 896
979 GlyPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 995
897 GCTTTGGTGTACAGCGCGCAAGCTCTCCCTGACAAATGTTATACC 946
995 nProProMetGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1010
947 AACCAACAGTGAATTTGTCGAGCTCCGATGCTG..... 981
1010 GlnGlnGlnGlnAsnMetMetAsnGlnProMetGlnSerAsnAsnGln 1026
982 .....GTGCAGCAGCCCGAGTGA 1001
1027 AsnMetAlaAsnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1043
1002 GCCCGAGGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1050
1043 eGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1058
1050 ..... 1050
1058 rOlyProArgGlyArgGlyLysLysAsnProAlaGlnGlnProProPro 1074
1051 .....CAGATGTGCTCCCGAGTCCAGCTCCAGCTCCAGCAGCAGCT 1091
1075 ProLysGlnGlnThrAsnAlaGlnLysLysLysLysLysLysLysLys 1091
1092 CCGCTG.....CTGTCT 1105
1091 uProMetGlyLysLysGlyLysGlyLysValMetAlaProSerLeuThrT 1108
1106 GCCGTCACCGGCGCAGGTCAGACCCCGCAGTGATGCCCTCC 1155
1108 hrAspThrPheHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1122
1156 CCGCAGCGCTCCCGCAGCGCGCAGCCAGCTCAGCAGCAGCAGCAGCAG 1205
1123 .....ProAsnAsnAlaGlyProMetLysGlnGlnGlnGlnGln 1135
1206 CGTCAGCTTGCGCTCCCGCAGCTCCAGTGTCTCTCCCGCAGCGCT 1255
1135 nLysAlaGlyLysProProProAsnGlnGlnGlnGlnGlnGlnGlnGln 1152
1256 CA.....CCGACGCTCCCGCAGCGCAG 1278
1152 sValProLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1168
1279 GTGACGCGCGCGGACCCAGCAAGCTTCACTGCTCCCTCAGCTGACCTT 1328
1169 ValGlnAlaLysSerSerProAspLysSerGlnProSerThrSerLysG 1185
1329 AACACACCTGTGAACCCAGC.....TGTGTC 1357
1185 nArgProProAlaLysLysLysLysLysLysLysLysLysLysLysLys 1202
1358 TGAAGCCAGCTGGCTCCAGCAGCTGAGGAGCAGCAGTACTGAGCAAG 1407
```

```

1202 aLserHisValLysGlySerArgLysMetThrProGlnuTyrlleGlu... 1217
1408 CTGAGACGAGCTGTGGAATACATC...GAGCCCTGGCCGCGATGATCAA 1454
1218 .....TyrValLysGlnLleuAsnArgLysLeuMet 1228
1455 CAAGATGACAGACAGACGAGAAAAGAGAC.....CTGAGTAA 1498
1228 gGluLeuGluLys...AlaAspMetLysThrAspArgArgTyrlleSerV 1244
1499 TGAAGACCTTCTGTGACATCTGTGACAGACCCCTGAGACCGGTGCCCTG 1548
1244 aLArgLysLeuLeuAlaLeuLleGlnLlu..... 1253
1549 AAGACCTTGCAAAAGTGTGAGATCGCCCTGAGACAACTCAAGATGACAT 1598
1254 .....GluLysLeuThrLleLleProPro..... 1262
1599 GCGGTGCCCACTCCCGCCAGCGC 1620
1263 .ArgLleProMetProAlaArg 1269

seq_name: p1r2:T30160
seq_documentation_block:
Hypothetical protein C37A2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30160
R:Le, T.T.: Kemp, K.; Scheet, P.
Submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C37A2.
A:Reference number: Z20746
A:Accession: T30160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1234 <LEU>
A:Cross-references: EMBL:U97194; PIDN:AB52447.1; GSPDB:GN00019; CESP:C37A2.2
A:Experimental source: strain Bristol N2; clone C37A2
C:Genetics:
A:Gene: CESP:C37A2.2
A:Map position: 1
A:Introns: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

alignment_scores:
Quality: 381.00 Length: 666
Ratio: 1.443 Gaps: 31
Percent Similarity: 39.640 Percent Identity: 27.027

alignment_block:
US-09-668-119-1 x T30160 ..

Align seg 1/1 to: T30160 from: 1 to: 1234
211 GCTGAATTGGATGCGCTCCTCGGGCGCCGGA.....CAGTC 248
403 SerGlySerGlnProThrProValProGlyThrProGlnProGlnG 419
249 TCTG.....GGCGGATGGGTAGCTTGTGTCATGGAGACG 286
419 nLleThrProGlnProGlySerLeuGlyProMetGlySerLeuGlyPro 436
287 CAATGCTCTCTCAGAGCGCCCTCTGGACCTCGGGATGGCCCT 336
436 rOthAla.....ProProGlySerGlnProMetAsnPro 447
337 CACAGCATGGGTGTGTGTACGCACTCCAGACCGACGTGACACT 386
447 .....ATGGTGTGACGACGCC..... 447
387 CCAGCAGGTGGCGCTGACAGACGACGACCAACAGCAGATTCCAGCAGC 436
|||||

```

```

448 .....GlnGlnLnaArgLleGlnG 455
437 AGCAGACGGCGCG..... 450
455 LngLngLnaLalProSerAlaSerAsnSerProLeuLeuValAsnLeu 471
451 CTACAGACACAGCAGCAGCAGCAGCAGCAGCAGTTC.....CA 491
472 LeuSerAsnGlnGlnProProGlnGlnGlnGlnGlnGlnGlnGlnGln 488
492 GGCCTCAGCAGAGTGCCTCAGCAGCAGCAGCAGTTCACAGCAGTGCAGCAGC 541
488 rAlaGlnGlnLysLeuSerMetGlnGln.....IleAlaLalLleGlnGlnG 503
542 AGCAGACAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTATTAATG 591
503 LngLln..... 505
592 CATCATCAAAATCAGCAACAGATACAGCAGCAGCAGCAGCAGCAGCAGCAG 641
506 HisGlnGlnTyrlGlnGlnLnaArgLleLeuGlnGlnGlnGlnGln..... 519
642 AATAGCAGAGCTGCAGCTCCAAACACAGCAGCAGCAGCAGCAGCAGCAGC 691
520 .....GlnAlaMetMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 530
692 AGCAGCAGCAGCAGCAGCAGCT..... 711
530 aLgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 546
712 .....TTGGAGGCCAGCAGCAGCA..... 729
547 ProGlnGlnPhePheProThrAsnGlnGlnAlaGlyProProGlyThrProG 563
730 .....ATTACAGCAGCCAGCAGT..... 747
563 YArgProLleProProTyrlAlaMetGlyLlnProPrometLysHisGlnA 580
747 ..... 747
580 lAcGlyProGlnGlnMetLleGlnArgMetAsnSerTyrlProGlyAsn 596
748 .....CAGCAGCAGCAGCAGCTCCGCTCCAGCAGT.....CTGCCCA 785
597 AlaGlnGlnPheArgProProProGlnGlnGlnGlnGlnGlnGlnGlnGln 613
786 GCAGCTGCAGCAGATGCATCACACAGCAGCAGCAGCAGCAGCAGCAGCAGC 835
613 oGlnGlnGlnGln.....GlnProProAlaProProGlnGlnL 626
836 CCCAGCAGCCTCCAGTGTGTCAGACCAACCAACCAACCAACCAACCAACCA 882
626 euGlnGlnProProValGlnAspThrSerAlaValAlaGlnProProLys 642
882 ..... 882
643 LysLysLysArgProThrLysLysGlnLysGlnAlaAlaAlaAlaAlaAla 659
883 .....CAGTCCAGACACCGAGCTTGTGTCACAGCGCCAGAGCTCTCCCTG 928
659 aLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 676
929 GACAAATGTTGTATACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 975
676 LysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 692
976 .....ATGGTGTGACGACGCC..... 993
693 GlnGlnGlnGlnMetValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 709
993 ..... 993
709 oGlyMetProProProGlnGlnLysAlaPheProProGlyTyrlLeuProGlyG 726

```

```

994 ..... CCAAGTCAGCCCGCCAGGTCAGCAGCAGCAGCAGCAGTACAG 1035
726 LVALGlnProthrGlnGlnInlterpGlnGlnGlnGlnGln 742
1036 ACAGCTCAGGCTGCC ..... CAGATGGGCTCCCGAGTCCAGGT 1076
743 GlnGlnMetAlaMetPheGlnGlnGlnGlnGlnGlnGln 759
1077 CAGCCAGAGCAGCCTCCCATG ..... CTGT 1102
759 GGLyLysProGlnGlnProMetGlyGlnTrpProGlnGlnArgLysLeuP 776
1103 CCTCCGCTCAGCCGCGCCAGCAGGTGACAGCCCGCAGTCCAGT ..... 1146
776 rovalProLysProProGlyLysAsnSerThrProAspAlaValAsnGln 792
1146 ..... 1146
793 GlnGlnAsnProLysProGlyAlaThrMetGlnHisArgLysSerGly 809
1147 ..... CCCCCTCC ..... CCCCAGCCGCTCC 1168
809 uPheAlaProProPrcValSerSerGlyLysValHisProGlnGlnGln 826
1169 CGCAGCCGCGCAGCCAGCAGTCACAGCCAGCAGTCCAGTCCAGTCCG 1218
826 InGlnGlySerPheHisArgSerSerSerAlaSerValTrpSerGly 842
1219 CCTGCCCATCTCCG/GTACGCTTCTGCTCCAGCCCTCAGCCAGCCCTC 1268
843 ..... SerHisThrProPheGlyGlnGlnGlnGlyProse 853
1269 CCAGAGCCAGTACGCGCGGACCCAGACAGACTTACGTGCTCCCTCAC 1318
853 rGlnSerGlnProThrAla ..... ValPro...P 862
1319 CTGACCTTTAAACACACCTGTGAACCCAGCCTGTATAGCCAGCAGT 1368
862 roGlyProGlnGlnAsn ..... AsnProGly 870
1369 GGCCTCCAGCCAGGCTCAGAGCAGCAGTACTGACAGCAGTCCAGT 1418
871 GlySerGlyAspIleGlyGlnLysAlaLysValAspGlnLeuAsn 887
1419 GTGCAAGTACATGACCCCTGCG ..... 1442
887 rSer ..... GluProLeuAlaAspLeuGlyAspLeuGlyAspLeuG 901
1443 ..... CCGCATGATACACAGATGACAGAGAGAGAGA 1475
901 LysAspLeuAspIleGlnProMetAspValGlnAspGlyGlnTrpProSer 917
1476 CAGAAAAAAGAGCCTGAGTAAGAGAGCCTTCTGAGACATTTCTGACAG 1525
918 ThrSerAsnGly ..... GluArgAsnGlnArgSerAspArg 929
1526 A ..... CCGCTCGAAGGCGTGTCCCTGAA ..... 1550
929 glLeuAlaLysSerIleGlnLeuValLysGlnValAlaLysSerGlyArg 946
1551 ..... GACCTTCAGAAAGTGGAGATGGCCCTGAGGAA 1583
946 laGlyAlaPheAsnAlaGlnLeuAlaLysIleGlyLysArgGly... 961
1584 ACTGAAGAATGACATGCGTGGTCCACATCCCGCCAGCCCGCGGTGCCA 1631
962 ..... SerValProLysProLeuProAlaLysSerPro 972

seq_name: p1r2:113606
seq_documentation_block:
hypothetical protein 87Bl.5 - fruit fly (Drosophila melanogaster)

```

```

C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13606; S23632
R:Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217668
A:Accession: T13606
A>Status: preliminary; translated from GR/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-1589 <MUR>
A:Cross-References: EMBL:Z98269; NID:e1355202; PID:e1251078; PIDN:CA10975.1
R:Decamilli, M.; Cheng, N.; Plerre, D.; Brock, H.W.
Genes Dev. 6, 223-232, 1992
A:Title: The polyhomeotic gene of Drosophila encodes a chromatin protein that shares
A:Reference number: S23632; MUID:92146957
A:Accession: S23632
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1010, 'V', 1012-1192, 'V', 1194-1274, 'T', 1276-1589 <DEC>
A:Cross-References: EMBL:X63672; NID:g11056; PIDN:CAA45211.1; PID:g11057
C:Genetics:
A:Gene: FlyBase:ph-P
A:Cross-References: FlyBase:FBgn0004861; FlyBase:FBgn0004860
A:Introns: 12/2; 595/1; 745/2; 1340/1
C:Superfamily: SAM homology
C:Keywords: DNA binding; nucleus
E:74-80, 247-285, 411-450, 494-650, 727-737, 775-955, 1032-1061/Region: glutamine-rich
E:1510-1576/Domain: SAM homology <SAM>

alignment_scores:
Quality: 378.00 Length: 629
Ratio: 1.24 Gaps: 32
Percent Similarity: 48.331 Percent Identity: 29.571

alignment_block:
US-09-668-119-1 x T13606 ..
Align seg 1/1 to: T13606 from: 1 to: 1589

19 GCACACAGTAAATCCAGCAGAGATATGAGACCATGTTTCTGAAGC 68
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
674 AlasSerValSerThrGlnThrAlaGlnAsnGlnSerLeuLysAl 690
69 CAGACCCGCGAGCAATAC ..... 87
111 111 111 111 111 111 111 111 111 111 111 111 111 111
690 alysmetArgAsnLysGlnGlnProValArgProAlaLeuAlaThrLeu 707
88 ..... CTTTCTCTCGTGCC ..... AGCCTCATTAATCCATT 120
111 111 111 111 111 111 111 111 111 111 111 111 111 111
707 ysthrGlnIleGlyGlnValAlaGlyGlnAsnLysValValGlyHisLeu 723
121 CGAGCATTCATACAGAAATCTCAGCTCCGAGTCCGATCCATGAA 170
111 111 111 111 111 111 111 111 111 111 111 111 111 111
724 ThrThrValGlnGlnGlnGlnGlnAlaThrAsnLeuGlnGlnValValAs 740
171 TGCAGTCCAGACCTGAGTGGCGGACCTGCGGAGCCGCTGGGAATG 220
111 111 111 111 111 111 111 111 111 111 111 111 111 111
740 n ..... AlaIleGlyAsnL 745
221 GCATGCTCTCTGGGGCCCGGAGCAGTCTGGCGGGATGGATGCTTT 270
111 111 111 111 111 111 111 111 111 111 111 111 111 111
745 ysmet ..... 746
271 GGTGCATGGGACAGCAATGTCTCTCAGGCGACCGCTCTGGGAC 320
111 111 111 111 111 111 111 111 111 111 111 111 111 111
747 ..... ValValMetSerThrThrGlyThrProIleThrLeuG 759
321 CTGGGGAATGGCCCTCAGACAGATGCTGTGCTGAGCGCACTCCAC 370
111 111 111 111 111 111 111 111 111 111 111 111 111 111
759 nAsnGlyGlnThrLeuHisAlaIleThrAlaAlaGlyValAspLysGln 776

```


1 . . . ,


```

      :||| :||| :||| :||| :|||
890 rgluglualagluProAlaIalaserArgaspAlaLysGlnGlnAspG 907
1573 GCCCTGGAGAACT 1586
      :||| :|||
907 luthrAlaIalThr 911

```

seq_name: swissprot_39:GBF_DICDI

seq_documentation_block:

```

ID   GBF_DICDI          STANDARD:          PRT:          708 AA.
AC   P36417;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   G-BOX BINDING FACTOR (GBF).
GN   GBFA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=AX3;
RX   MEDLINE=94170994; PubMed=8125261;
RA   Schmitzler G.R., Fischer W.H., Firtel R.A.;
RT   "Cloning and characterization of the G-box binding factor, an
RT   essential component of the developmental switch between early and
RT   late development in Dictyostelium."
RL   genes Dev. 8:502-514(1994).
CC   1- FUNCTION: CAMP-RESPONSIVE TRANSCRIPTIONAL ACTIVATOR REGULATING
CC   LATE GENE EXPRESSION. ESSENTIAL COMPONENT OF THE DEVELOPMENTAL
CC   SWITCH BETWEEN EARLY AND LATE DEVELOPMENT. BINDS TO A NUMBER OF
CC   CA/GT-RICH GENE REGULATORY ELEMENTS.
CC   -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

```

```

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

```

```

DR   EMBL; L29075; AAA21021.1; -.
DR   TRASNFAIC; T00315; -.
KW   DictyDdb; DD02046; gbFA.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Repeat.
FT   REPEAT          339..368             1.
FT   REPEAT          481..510             2.
FT   DOMAIN          11..21               POLY-SER.
FT   DOMAIN          115..263             GLN-RICH.
FT   DOMAIN          270..292             POLY-ASN.
FT   DOMAIN          549..557             POLY-ASN.
SQ   SEQUENCE 708 AA; 79268 MW; B4B6D8F04FACACCA CRC64;

```

alignment_scores:

Quality:	383.00	Length:	697
Ratio:	1.272	Gaps:	23
Percent Similarity:	43.185	Percent Identity:	24.390

alignment_block:

US-09-668-119-1 x GBF_DICDI ..

Align seg 1/1 to: GBF_DICDI from: 1 to: 708

```

130 CATACAGAAATCTCAAGCTCGTCAGATGATCTATGATGACATCA 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6 HishisglnGlnYanSerSerSerSerSerSerSerSerProsergl 22
180 GAGGCTGAGTGGCGGACCTGCTGGGAGCGCGCTGAATTGGCATGGCT 229
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

22 nThrIleGlySerAspLeuSerAsnIleSerAlaLeuProLeuProL 39
230 CTCGGGGGC..... 237
39 euProSerIlePheThrAlaGlnAsnGlnMetAsnProIleLeu 55
238 .....CCGGACAGTCT...CTGGCGGAGTGGGTAGCTTTGGTGGCAT 278
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
56 pheProProThrSerSerLeuLeuGlySerSerAsnThrProSerPh 72
72 GGGACAGCCATGCTCTCTCAAGGCGAGCCGCTCTGGGACCT 322
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
323 CGGGATGGCCCTTCACAGCATGGCTGCTGCTCT..... 357
89 spGlyGlnIleProAspMetProAsnMetValAspGlnIleLeuHis 105
358 ...ACGGCAACTCCACAGACCAGCTGCAGCTCCAGAGTGGCGCTTCA 404
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
106 ProAsnGlnAsnProHisIleTyranIleGlnIleLeuMetPheMetGl 122
405 GCAGCAGCAGCAGCAGCAGCAGCTTCCAGCAGCAGCAGCAGCGGCGCTAC 454
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
122 nGlnGlnAlaGlnGlnAsnGlnProProGlnGlnAsnGln.....G 136
455 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTCCAGCTCAGCAGAGT 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
136 nGlnGlnIleHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 152
505 GCCATCAGCAGCAGCAGCTTCCAGCAGTATGTCAGCAGCAGCAGCTTCA 554
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 MetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHis 169
555 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTAAATTCATATCAATTC 604
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
169 sGlnGlnMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 186
605 AGCAGCAGCAGTACAG.....CAGCAGCAGCAGCAGCTGAGGAGTAA 645
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
186 euGlnGlnIleHisIleGlnIleGlnGlnGlnGlnGlnGlnGlnGlnHis 202
646 GCAGCAGCTCAGCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 695
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
203 GlnGlnGlnGlnHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnHis 219
696 GCAGCAGCAGCAGCTTGGAGGCCAGCCAGCAGCAATTCAGCAGCAGCA 745
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219 nGlnGlnGlnGlnHisIleGlnIleSerGlnProGlnGlnGlnHis 236
746 TGCAGCAGCAGCAGCTCCGCGCTCCAGGCTCTGCCAGCAGCTGAG 795
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
236 sGlnGlnGlnGln.....HisGlnHisGlnGln 245
796 CAGATTCATCAGCAGCAGCAGCAGCAGCCGCGCAGCAGCAGCAGCAGC 845
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 GlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPro 262
846 T..... 846
262 oGlnSerLeuSerAsnSerGlyAsnAsnAsnAsnAsnAsnAsnAsnA 279
846 ..... 846
279 snAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnSerHisGln 295
847 .....CAAGTTCCTCAGAACCAACCATCA..... 870
296 LeuAsnAsnLeuThrLeuSerGlnAsnAsnThrSerGlySerAsnThrPr 312
870 ..... 870
312 oSerProSerThrIleGlyGlyArgGlyHisIleGlnIleThrSerAsnSer 329

```

```

1
870 ..... 870
329 LulysLysaspserSerGlyGlnThrLleProLysCysThrArgCysAsn 345
870 ..... 870
346 GluAlaIaIaSerTrpLysHisAspLysArgArgTrpTrpCysLysGluCys 362
871 .....CAACTCCGCCAC 883
362 sLysLysAlaPheThrProGlyLleThrLysMetGlnGlnValProGln 379
884 AGTCGACAGACCCAGCCCTTGTGTGCACAGGCGCAGACTCTCCCTGCAGAA 933
379 LulAgLlLeuGlnPrLleuGlnAsnHisAsnGlnLlePro...Gln 394
934 ATGTTGTATACCCACACACCACTGAATTTGTCCGAGCTCCGATGTGTGT 983
395 LeuTrpAspSerGln.....GlnAsnAsnSerSerGln 405
984 GCAGCAGCCCGCCAGTCAGCCCGCAG..... 1008
405 nasThrProProThrGlnProGlnAsnAsnMetAsnGlnLleAsnHisG 422
1009 .....GTGACAGAGCAG.....CAGACAGCAGTACAG 1035
422 LulLeuGlnGlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 438
1036 ACAGCTCAGAGCTGCGCCAGATGTGTGTGTGTGTGTGTGTGTGTGTGT 1074
439 LeuThrAlaSerAsnGlnGlnValProProLlLeuGlnGlnGlnLleAs 455
1075 .....GTGACCCAGAGCAGCCTCC 1093
455 nGlyLlLeuProAsrAsnAsnAsnSerLeuLleThrGlnAsnThrLeuA 472
1094 CCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1143
472 snSerLeuSer.....ThsSerValSer 479
1144 ATGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1158
480 CysProProCysProLleuGlnArgLylleSerSerTrpLysHisAsp 496
1158 ..... 1158
496 sLysArgTrpPheCysLysGlnCysLysLysLysProPheThrProValGlyA 513
1159 .....CAGCCGTCGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1200
513 lAgLylLeuSerProSerSerSerProSerProLysLysLysSerAsn 529
1201 .....TCCACGTCAGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1229
530 lIleThrProLlLeuThrThrSerSerThrSerSerSerSerThrProSe 546
1230 T..... 1230
546 rLlelIleAsnAsnAsnAsnAsnAsnAlaAsnSerSerLysAsnAsn 563
1231 .....CCCATGACCTCTGCTGCGCCAGCCCTCA 1257
563 hrProLysLysGlnLeuSerProProProSerValLeuGlnSerProSer 579
1258 CCGCAGCGCC...TCCGAGAGCCAGTG.....ACGGCGCG 1289
580 SerSerSerLleSerGlnSerProLeuGlnLeuAsnTrpGlnThrProth 596
1290 GACCCACAGACTTCACTGCTCCCTCA...CCTGACCTTTAAACAC 1336
596 rCysSerProAsnProSerLleuProSerLlGlyLysAsnLeuAsnSerG 613

```

```

1337 CTGNGAAC.....CCGAGCTGTCTATGAGCCCGAGTGGCTCCAGC 1377
613 LulAlaAsnSerLlLeuProAspGlyLylleuLleSerGlyLeuSer 629
1378 CAGCTGAGGAGCAGCAGTACCTGACACAGCTCAGACAGCTGTCCAG.. 1425
630 ProProLysSerSerSerSerLleAsnAsnLysSerPheSerAsnTh 646
1426 .....TACATGAGCCCTGCGCGCA 1447
646 rGlyAlaLleuLeuSerSerAsnGlyLlLeuAsnLleAlaAsnLeuGlyAsn 663
1448 TGATCAACAAGATGACACAGACGAGAGAGAGAGAGAGAGAGAGAGAG 1488
663 rLleuSerGlnLeuAsnLysGlnLysLysArgSerAsp 676

seq_name: SwissProt_39:MAM_DROME

seq_documentation_block:
ID MAM_DROME STANDARD: PRI: 1596 AA.
AC P21519:
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE NEUROGENIC PROTEIN MASTERMIND.
GN MAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RC MEDLINE=9106516; PubMed=1701150;
RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA Vedvobnick B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers."
RL Genes Dev. 4:1688-1700(1990).
CC - FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC - SUBCELLULAR LOCATION: NUCLEAR.
CC - DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBQUITOUSLY LOCATED, DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC - MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
CC - SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54251; CAA38152.1; -.
DR PIR: A3106; A3106.
DR PIR: A36391; A36391.
DR FLYBase: FBgn0002643; mam.
KW Neurogenesis; Nuclear protein; Repeat.
FT DOMAIN 20 84 GLN-RICH.
FT DOMAIN 127 190 ARG/LYS-RICH (BASIC).
FT DOMAIN 196 219 GLN-RICH.
FT DOMAIN 259 304 ASN-RICH.
FT DOMAIN 355 388 GLY/ASN-RICH.
FT DOMAIN 392 406 GLN-RICH.
FT DOMAIN 407 440 GLY-RICH.

```



```

768 erSerSerThrGluSerSerSerAlaProValProThrProSerSer 784
1563 AACCTAGAGATGACATCGGTCGCCACACCCGCCGCCGCTGCCAC 1632
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
785 SerThrThrGluSerSerSerAlaProValProThrProSerSer 801
1633 CGACCAACAGACAGTA...CCTATGCCA..... 1657
      | ::::: ||| ||| ||| ||| ||| ||| ||| |||
801 ThrGluSerSerValAlaProValProThrProSerSerSerAsnI 818
1658 .....GCCGCTCTGGATGCCCTGCGCAACATCCGCTCACCTGT 1699
      ::||| ||| ||| ||| ||| ||| ||| ||| |||
818 IeThrSerSerAlaProSerSerThrProPheSerSer...ThrIu 833
1700 CTTCAACCATTCCTGCTGACGACATTCGTCGACG 1735
      ::||| ||| ||| ||| ||| ||| ||| ||| |||
834 SerSerSerValProAlProThrProSerSerSer 845
seq_name: SwissProt_39:ANP_NOTCO

```

```

seq_documentation_block:
ID ANP_NOTCO STANDARD; PRT; 790 AA.
AC P24856;
DT 01-MAR-1992 (Rel. 21, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANTIREEZE GLYCOPETIDE POLYPROTEIN PRECURSOR (AFGP POLYPROTEIN)
DE [CONTAINS: AFGP7 (AFGP 7); AFGP8 (AFGP 8)] (FRAGMENT).
GN AFGP8.
OS Notothenia coriiceps neglecta (Black rockcod) (Yellowbelly rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Notoleosteii;
OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;
OC Nototheniidae; Nototheniidae; Notothenia.
OX NCBI_Taxid=8209;
RN [1]
RX MEDLINE=91067687; Pubmed=2251271;
RA Hsiao K.-C., Cheng C.-H.C., Fernandes I.E., Detrich H.W. III,
RA Devries A.L.;
RT "An antifeeze glycopeptide gene from the antarctic cod Notothenia
RT coriiceps neglecta encodes a polypeptide of high peptide copy
RT number."
RT Proc. Natl. Acad. Sci. U.S.A. 87:9265-9269(1990).
RN [2]
RP REVISIONS TO N-TERMINUS AND 457.
RA Cheng C.-H.C.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ANTIREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED BY THE LIVER AND SECRETED INTO
CC THE BLOOD FROM WHICH THEY BECOME DISTRIBUTED TO ALMOST THE ENTIRE
CC EXTRACELLULAR SPACE.
CC -1- DOMAIN: CONTAINS 44 COPIES OF AFGP8 AND TWO COPIES OF AFGP7.
CC -1- PTM: THE DISACCHARIDE GALACTOSE-N-ACETYLGALACTOSAMINE IS ATTACHED
CC TO THREONINES IN AFGP8 AND AFGP7.
CC -1- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 5 of December 2000;
CC WWW=http://www.expasy.org/spotlight/articles/sptl005.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; M55000; AAA49392.2; .
DR PIR; A38420; A38420.
DR HSSP; P04002; IWFA.
DR InterPro; IPR00104; Antifeeze_1.

```

```

DR InterPro; IPR001778; POA_allergen.
KW Antifeeze protein; Glycoprotein; Polypeptide; Repeat;
KW Multigene family.
FT NON_TER 1
FT PROPEP 1
FT CHAIN 6
FT PEPTIDE 6
FT PEPTIDE 23
FT PEPTIDE 40
FT PEPTIDE 57
FT PEPTIDE 74
FT PEPTIDE 91
FT PEPTIDE 108
FT PEPTIDE 125
FT PEPTIDE 142
FT PEPTIDE 159
FT PEPTIDE 176
FT PEPTIDE 193
FT PEPTIDE 210
FT PEPTIDE 227
FT PEPTIDE 244
FT PEPTIDE 261
FT PEPTIDE 278
FT PEPTIDE 295
FT PEPTIDE 312
FT PEPTIDE 329
FT PEPTIDE 346
FT PEPTIDE 363
FT PEPTIDE 380
FT PEPTIDE 397
FT PEPTIDE 414
FT PEPTIDE 431
FT PEPTIDE 448
FT PEPTIDE 465
FT PEPTIDE 482
FT PEPTIDE 499
FT PEPTIDE 516
FT PEPTIDE 533
FT PEPTIDE 550
FT PEPTIDE 567
FT PEPTIDE 584
FT PEPTIDE 597
FT PEPTIDE 614
FT PEPTIDE 631
FT PEPTIDE 648
FT PEPTIDE 665
FT PEPTIDE 682
FT PEPTIDE 699
FT PEPTIDE 716
FT PEPTIDE 736
FT PEPTIDE 756
FT PEPTIDE 773
FT PEPTIDE 790
SO SEQUENCE 790 AA; 71265 MW; 4CYCCACDC48FE902 CRC64;

```

```

alignment_scores:
Quality: 330.00 Length: 398
Ratio: 1.460 Gaps: 13
Percent Similarity: 56.784 Percent Identity: 33.166

```

```

alignment_block:
US-09-668-119-1 x ANP_NOTCO ..

```

```

Align seg 1/1 to: ANP_NOTCO from: 1 to: 790
285 GCCAATGCTCTCTCAGAGGACCGCCCTCTGGACCTGGGGATGCCC 334
      |||:::||||| ||| ||| ||| ||| ||| ||| |||
220 AlaThrAlaLeuAsnPhaAlaThrAlaAlaThrAlaAlaThrPr 236
335 CTCACACATGCTGCTGCTACGCACTCAGACCA..... 377
      |||:::||||| ||| ||| ||| ||| ||| ||| |||
236 AlaThrAlaAlaCysAsnPhaAlaAlaThrAlaAlaThrProAlaThra 253

```



```

|||||
829 pleuValHisLysThrProGlnProIleuAspGlnIleAsnGlyThrG 846
1532 CGAAGCGGTGTCCTCCGAGACCTTGCAAGAGT..... 1566
846 IlystLysLysLysLeuSerLysAlaAlaGlnLysAlaArgAspGlnAspPro 862
1567 ...GAGATCCGCTG.....CA 1580
863 ValGlnIleSerValAlaArgAsnLysLeuLeuMetProSerLysSerGI 879
1581 GAAGCTCAGATGACATGCGGTGCGCCACCTCCCGACCGCCCGCGTGGC 1630
879 ulYstThrLeuArgSerPheLysIleProIleAlaAspIleThrAlaCysP 896
1631 ACGGACCA 1638
896 heLysPro 898

```

seq_name: SwissProt_39:FSH_DROME

```

seq_documentation_block:
ID FSH_DROME STANDARD: PRT: 2038 AA.
AC P13709; P13710;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE FEMALE STERILE HOMEOTIC PROTEIN (FRAGILE-CHORION MEMBRANE PROTEIN).
GN FSH1H OR FSH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89276730; PubMed=2567251;
RA Haynes S.R., Mozer B.A., Bhatia-Bey N., David I.B.;
RT "The Drosophila fsh locus, a maternal effect homeotic gene, encodes
RT apparent membrane proteins."
RL Dev. Biol. 134:246-257(1989).
CC - FUNCTION: REQUIRED MATERNALLY FOR PROPER EXPRESSION OF OTHER
CC HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC - SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC - SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC - SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M23221; AAA28540.1; -
DR EMBL: M23222; AAA28541.1; ALT_TERM.
DR EMBL: M15762; AAA70424.1; -
DR EMBL: M15763; AAA70423.1; -
DR EMBL: M15764; AAA70422.1; -
DR PIR: A43742; A43742.
DR HSSP: P04002; 1WFA.
DR FLYBase: FBgn0004656; fsh(1)h.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain_2.
DR PRINTS: PR00503; bromodomain_2.
DR SMART: SM00297; BROMO_2.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS50014; BROMODOMAIN_2; 2.
KW Developmental protein; Bromodomain; Transmembrane; Repeat.
FT DOMAIN 51 123 BROMODOMAIN 1.
FT DOMAIN 495 567 BROMODOMAIN 2.
FT DOMAIN 945 1106 ET DOMAIN.

```

```

FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 750 770 POTENTIAL.
FT TRANSMEM 790 810 POTENTIAL.
FT TRANSMEM 816 830 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
FT TRANSMEM 1731 1751 POTENTIAL.
FT TRANSMEM 1939 1959 POTENTIAL.
FT VARIANT 909 909 G -> A.
FT VARIANT 1022 1022 H -> RRPVY.
SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

```

alignment_scores: Quality: 327.50 Length: 559
Ratio: 1.269 Gaps: 20
Percent Similarity: 46.154 Percent Identity: 26.834

alignment_block:

US-09-668-119-1 x FSH_DROME ..

Align seg 1/1 to: FSH_DROME from: 1 to: 2038

```

130 CATACAGAGAA.....TCGACAGCTTCGTCAGTGATCC 164
|||||
1226 HISAsnLysAsnGlyProAsnAspLeuSerValGlnProGlyGlyPr 1242
165 TATGATGACACTC.....CAGAGCCTGATGCGGAGCCTGTCGGG 205
|||||
1242 oLLeSnAlaAlaLeuProProHisSerPheAlaGlyThrAlaThrV 1259
206 GAGCC..... 210
1259 aLAlaThrSerGlnSerSerGlyGlyLeuArgIleAlaSerAsnLeuHs 1275
211 .....GCTGGAATTGGCATGCTCTCTCGGGCGGAGCATGTCTGGG 254
1276 LysProSerGlyLeuGly.....GlyLysAspLeuGI 1286
255 C.....GGGATGGGTAGCT 268
1286 YGlnHisHisAlaAlaLeuAlaAlaAlaLeuThrSerGILeuSert 1303
269 TTGGTCCATGGGACAGCCATGTCTCTCAGG..... 303
1303 hrgGlyThrAlaGlyGlyGlyLeuAsnAsnGlnGlySerAsnAsn 1319
304 ...CAGCCGCTCTGGGACCTCGGGATGGCCCTCAGACATGGCTGT 350
|||||
1320 AlaAsnProLeuGlyGlySerHisGlyAspAlaMetValAsnAlaSerLe 1336
351 CGTGCTAGGCGCAGCTCCAGACCCAGCTGCAGCTCCAGACAGTGGCGC 400
1336 uAlaSerLeuAlaSerGlyLeuLysGlnIleProGlnPheAspAspProV 1353
401 TGCAGCAGCAGCAGCAGCAGCAGCAGTTC..... 429
|||||
1353 aLGIuGlnSerLeuAlaSerLeuGlnPheSerAlaGlySerThrGlyLys 1369
430 .....CAGCAG.....CAGCAGCAGGC 446
1370 SerGlyLeuThrAspAsnPheLeuMetGlnGlnHisLeuMetGlnProAl 1386
447 GGGCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTTCAGGCTC 496
1386 aGlyProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1403
497 AGCAGAGTGCATGCAGCAGCAGCAGTTCAGCAGTACGTAGTCAGCAGAGC 546
|||||
1403 GlnGlnGln.....GlnGlnGlnGln 1409
547 CAGCTCAGCAGCAGCAGCAGCAGCAGCAGCTA.....ATTAA 587
|||||

```


Ratio: 1.26 Gaps: 35
Percent Similarity: 48.370 Percent Identity: 30.703

Alignment block:

US-09-668-119-1 x GLUT4_WHEAT

Align seg 1/1 to: GLUT4_WHEAT from: 1 to: 838

```

39  GGATATGGAGAGCCAGTGTTCCT...GAGGCCAAGACCCGGAGCAT 85
   |||.....|||
126  GlnGlnAlaSerProGlnArgProGlnGlnGlnGlnProGlnGlnGln 142
   |||.....|||
86  ACCTTCTCT...CGTGGC...CAGGCTCAATATTCATTTGAGACATT 129
   |||.....|||
142  yGlnGlnSerGlyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 157
   |||.....|||
130  CATACAGAAATCTCAAGCTCCGTCAAGTATCTATGATGCACTCA 179
   |||.....|||
158  ....GlnGlnProGlnGlnProGlnGln.....Pro 166
180  GAG.....CCTCACTGGGAGCCTGGCTGGCGGAGCCGCTGGAATTG 220
   |||.....|||
167  GlnGlnGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 177
   |||.....|||
221  GCATGCTCTCTGGGCGCCGGAGAG.....TCCTGGGGGG 258
   |||.....|||
178  ....ProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 191
   |||.....|||
259  ATGGGTAGCTTGGTGGCATGGGAGCAGCAATGCTCTCTCAGGCGAGCC 308
   |||.....|||
192  GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 202
   |||.....|||
309  GCCCTCTGGACCTCGSGGATGGCCCTCAAGCAGTGGCTGTGCTTA 358
   |||.....|||
202  oglyGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 213
359  CGGCACTCCACAGACCCAGCTG.....CAGCTCCAGCAGTGGCG 359
   |||.....|||
214  ....SerGlnLeuGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnProAla 225
400  CTG...CAGCAGCAGCAGCAGCAGCAGCAGTCCAGCAGCAGCAGCAGC 446
   |||.....|||
226  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 242
   |||.....|||
447  GGGCCTACAGCAGCAG.....CAGCAGCAGCAGCAGCAGCAGCAGTTC 490
   |||.....|||
242  oglyGlnGlnGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 259
491  AGGCTCAGCAGAGTGCATGCAGCAGCAGTCCAGCAGCAGTGCAGCAG 540
   |||.....|||
259  GlnGlnGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 275
541  CAGCAG.....CAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCA 575
   |||.....|||
276  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 292
   |||.....|||
576  GCATCTAATTAATGATCATCAAAATCAGCAACAGATACAG..... 618
   |||.....|||
292  oglyTyrTyrProThrSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 309
   |||.....|||
619  ....CAGCAGCAGCAGCAGCAG.....CAGCAGCAGCAGCAGCAG 633
   |||.....|||
309  yTyrProThrSerProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGln 325
   |||.....|||
634  CTGCAGCAGATAGCAGAGTGCAGCTCCAGCAACAGCAGCAGCAGCAGCA 683
   |||.....|||
326  LeuGlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 342
   |||.....|||
684  GCAG.....CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 727
   |||.....|||
342  nGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
   |||.....|||
728  CATTCAG...CAGCA.....CCGATGCAG..... 750

```

```

359  roGlyGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 375
   |||.....|||
751  GAGCCACAGCCT.....CCGCCCTCCAGGCTCTGCC..... 783
   |||.....|||
376  GlnGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 392
   |||.....|||
784  ....CAGCAGCTCCAGCAGATGATCACAACAGC 814
   |||.....|||
392  rGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 409
   |||.....|||
815  ACCACAGCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 864
   |||.....|||
409  AlaGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 423
   |||.....|||
865  CCATCACAACCTCCGCCACAGTCCGAGACCCAGCCTTGTGTCTCAGAGC 914
   |||.....|||
424  ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 964
   |||.....|||
915  GCAAGCTCTCCCTGGACAAATGTTGATACCAACCAACCATCAAAATTG 964
   |||.....|||
435  yGlnGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 443
   |||.....|||
965  TCCGAGCTCCAGTGTGTGTCAGACGCCCGCAGTCCAGCCCGAGTGCAG 1014
   |||.....|||
444  ....ThrSerProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 453
   |||.....|||
1015  CAGCAGCAGCAGCAGCAGTACAGACAGCAGTCCAGGCTGCCAGATGGTCTCC 1064
   |||.....|||
454  GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 470
   |||.....|||
1065  CGAGTCCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1114
   |||.....|||
470  ogly...GlnGlnGlnGln.....ProGlyG 478
   |||.....|||
1115  CGGCGCAGCAGTGCAGACCCCGCAGTCAAGTCCAGCAGCAGCAGCAGCAG 1164
   |||.....|||
478  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 494
   |||.....|||
1165  TCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1214
   |||.....|||
495  Gly.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
   |||.....|||
1215  TGGCCCTGCCATCTCCAGTACGCTCTCCAGCAGCAGCAGCAGCAGCAGCAG 1264
   |||.....|||
509  rGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 526
   |||.....|||
1265  CCTCCAGAGCCAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1302
   |||.....|||
526  roGlyTyrTyrProThrSerProLeuGlnProGlnGlnGlnGlnGlnProGly 542
   |||.....|||
1303  TTCAGTGTCCCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1352
   |||.....|||
543  TyrAspProThrSerProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGln 559
   |||.....|||
1353  TGTCTAGAGCCAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1395
   |||.....|||
559  nLeuGlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 576
   |||.....|||
1396  TACCTGCAGCAGCAGTGAAGCAGCAGTGCAGCAGCAGCAGCAGCAGCAGCAG 1445
   |||.....|||
576  GlnGlnGlnGlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnProAla 591
   |||.....|||
1446  CATGATCAACAAGATCGACAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1495
   |||.....|||
592  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 605
   |||.....|||
1496  AGATGAAGAGCCTTGTGACATTCGACAGACCCCTGGAAGCGGTCTCC 1545
   |||.....|||
606  ....GlnGlnPro...GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 620
   |||.....|||
1546  CTGAAGACCTTGCAAAAGTGTGATGATGCCCTGAGAGAACTCAAGAA 1592
   |||.....|||

```

620 yglIn.....GlnProGlyGlnGlyGln 628

seq_name: SwissProt_39:GLT5_WHEAT

seq_documentation_block:

ID GLT5_WHEAT STANDARD: PRT: 839 AA.

AC P10388;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT DX5 PRECURSOR.

GN GLU-1D-1D OR GLU-D1-1B.

OS Triticum aestivum (wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triticeae; Triticum.

OX NCBI_TaxID=4565;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. CHEYENNE;

RX MEDLINE=89098419; PubMed=2563152;

RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,

RA Malpica-Romero J.M.;

RT "Nucleotide sequences of the two high-molecular-weight glutenin genes

RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv

RT Cheyenne.";

RL Nucleic Acids Res. 17:461-462(1989).

RN [2]

RP REVISIONS, SEQUENCE FROM N.A.

RC STRAIN=CV. CHEYENNE;

RA Anderson O.D.;

RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE

CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE

CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.

CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.

CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE

CC GROUP 1 CHROMOSOMES OF WHEAT.

CC -1- MISCELLANEOUS: THE NATURE PROTEIN IS CHARACTERIZED BY A LARGE

CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GOQPEQ AND

CC GOQPEQGOQGYPTS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X12928; CAA31395.1; -

DR PIR; S02262; S02262.

DR InterPro: IPR001419; GLUTENIN.

DR PRINTS; PR00210; GLUTENIN.

KM Seed storage protein; Repeat; Multigene family; signal.

FT SIGNAL 1 21

FT CHAIN 22 839 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT

FT FT 22 839 DX5

FT REPEATS.

FT DOMAIN 131 801

FT SUBDOMAIN 839 AA; 89359 MW; 0F14E110D552643 CRC64;

SQ

alignment_scores:

Quality: 316.00

Ratio: 1.141

Percent Similarity: 45.041

Percent Identity: 28.780

alignment_block:

US-09-668-119-1 x GLT5_WHEAT

Align seg 1/1 to: GLT5_WHEAT from: 1 to: 839

39 GGATATGAGAGCCATGTTTCT...GAAGCCAGACCCGGAGCAAT 85

```

111  ::::: 111  :::::
127 gylGlnAlaserProGlnArgProGlyGlnGlnGlnProGlyGln. 142
86 ACCTTCTCTCTGTCGTCGACAGCTCATTCATTCATTTTCAGACAT..... 128
143 .....GlyGlnGlnGlyTyrTyrProThrSerProGlnGlnProG 156
129 ..TCATACAGAAATTCACAGCTTCGTCAGTGTCTATGATGACCT 176
156 lylGlnTrpAlaInProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 172
177 CCAGAGCTCAGCTGC.....GACCTGCTCGGGAGCCGCTGGA 217
173 ProGlnGln. ProGlyGlnLeuGlnGlnProAlaGlnly..... 185
218 TTGGCATCTCTCTCGGGCCCGGAGACGTCCTGGGGGATGGGTGC 267
186 .....GlnGlnProGlyGln..... 190
268 TTTGGTCCATGGAGACCATGTCCTCTCAGGGACGCCCTCTGG 317
191 ..GlyGlnGlnGln.....GlnProGlyGlnG 200
318 GACCTCGGGGATGACCCCTCAGACGATGGCTGTCTAGCGCACTC 367
200 yglInProGlyTyrTyrPro..... 206
368 CACAGACCCAGCTCAGCTCCAGCAGCTGCGCTCAGCAGCAGCA 417
207 .....ThrSerSerGlnLeuGlnProGlyGlnLeuGlnProAlaGln 221
418 CAGCAGCAGCTCCAGCAGCAGCAGCGGCGCTACAGCAGCAG..... 462
222 GlyGlnGlnGlnGlnProGlyGlnAlaGlnGlnGlnGlnProG 238
463 .....CAGCAGCAGCAGCAGCAGCAGCTTCAGG 493
238 yglGlnGlnGlnProGlyGlnGlnGlnGlnGlnGlnProGlyG 255
494 CTCAGCAGAGTGCATGCAGCAGCAGCTTCAGCAGCAGTATGAGCAGCAG 543
255 lylGlnGlnProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 271
544 CAG.....CAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 578
272 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProG 288
579 TCTAATTAAATGCATCATCAAAATCAGCAACAGATACG..... 618
288 YTYTYTYProThrSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 305
619 .....CAGCAGCAGCAGCAGCTG 636
305 ylrProThrSerProGlnGlnProGlyGlnGlnGlnGlnGlnGlnGln 321
637 CAGCGAATAGCAGAGCTCAGCTC.....CAACAACAGCAACAGCAGCA 680
322 GlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 338
681 GACAGCAGCAGCAGCAGCAGCAGCAGCAGCTTTGGAGCCAGCAGCAGCA 730
338 nProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProG 355
731 TTCAG...CAGCA.....CCGATGCAG.....CAG 753
355 lylGlnGlnGlnProGlyTyrTyrProThrSerProGlnGlnGlnGln 371
754 CCAGAGCT.....CGGCCCTCCAGGCTCTGCC..... 783
372 GlyGlnProGlyTyrTyrProThrSerSerGlnGlnProThrGlnSerG 388
784 .....CAGCAGCTCAGCAGCAGATGATCATCACACAGCAGCAGC 817

```



```

FT DOMAIN 302 305 POLY-PRO.
FT DOMAIN 376 382 POLY-SER.
FT DOMAIN 386 397 POLY-SER.
FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.
FT DOMAIN 704 707 POLY-PRO.
FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT CONFLICT 94 MISSING (IN REF. 2).
FT CONFLICT 333 Y -> H (IN REF. 2).
FT CONFLICT 339 M -> I (IN REF. 2).
FT CONFLICT 341 P -> T (IN REF. 3).
FT CONFLICT 541 G -> A (IN REF. 2).
FT CONFLICT 1028
SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

```

```

alignment_scores:
    quality: 316.50      Length: 637
    ratio: 0.995         Gaps: 38
Percent Similarity: 49.922 Percent Identity: 27.786

```

alignment_block:

US-09-668-119-1 x DRPL_HUMAN ..

Align seg 1/1 to: DRPL_HUMAN from: 1 to: 1185

```

163 CCTATGAATGACATCCAGAGCCTGACATGGGAGGAGCCTGCGGAGGAGCGC 212
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
303 ProProProAlaLeuAlaGlyProLeuAlaHisAla...SerAlaSerPro 318
213 TGAATTTGGATGCTCTCTCTGCGGCGCGGAGACATCTCTGCGGCGGATGG 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
318 oglyleuglyAlaGlnProLeu...ProGlyHis.....LeuPro 330
263 GTACCTTTGGTGGCATGGGAGCAGCATGCTCTCTGAGGAGCGCCCT 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 roSerProGlyAlaMetGlyGlnGlyMet.....GlyGlyLeuPro 343
313 CCTGGG.....ACCTGGGATGGCCCTCCACAGCATGGC 347
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 ProGlyProGlyGlyGlyProThiLeuAlaProSerProHisSerLeuPro 360
348 TGTGGTGTCTACGGCACTCCACAGACCCAGCTGACCTCCAGAGGTGG 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 oProAlaSerSerSerAlaPro.....PrometArgPhe 373
398 CGCTGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 ProTySerSerSerSerSerSerSerAlaAlaAlaSerSer..... 388
448 GCGGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
389 .....SerSerSerSerSerSerSerAlaSer. ProPheProAlaSer 402
498 GCAGAGTGC.....CATGCAGCAGCAGCTCCAGCAGCTAGTGC 535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 GlnAlaLeuProSerTyProHisSerPheProProProHisSerLeu 419
536 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 lValSerSngInProGlyTyThrGlnProSerLeuProSerGlnA 436
586 AATTCGATCA..... 596
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 lavalITPserGlnGlyProProProProProTyGlyArgLeuLeu 452
597 ...TCGAATTCGACAGATATCAGCAGCAGCAGCAGCAGCAGCAGCAG 643
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
453 AlaAsnSerAlaHisProGlyProPheProProSerThrGlyAlaG 469

```

```

644 TAGCAGCAGCTGCAGCTCCA.....ACACAG 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 nSerThrAlaHisProProValSerThrHisHisHisHisGln Gln 485
670 CAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTGGAGGC 719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
486 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisGlyAsnSe 502
720 CCAGCCCAACATTCAGACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 rGlyProPro.....ProProGlyAlaPheProHisProLeuGlnG 516
762 ..... 762
516 lGlySerSerHisHisAlaHisProTyAlaMetSerProSerLeuGly 532
763 .....CCGCCCTCCAGCT...CTGCCAGCAGCTGCA 794
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
795 GCAGATGCATCAGACAGACA.....CCAGCGCGCCAGCA 831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 rGlnValSerTySerGlnAlaGlyProAsnGlyProProValSerSerS 566
832 CAGCCCGCAGCAGCTCCAGCTTGTCTGAGAACCAACATCAGACATCCGCC 881
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 eSerAsnSerSerSerSerThrSerGlnGlySerTy.....ProGlyS 580
882 ACAGTGCAGCCAGCAGCTTGTGTCTGACAGCGCAGACAGCTCTCTCTGAGC 931
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
581 SerHisProSerProSerGlnGlyProGlnGlyAlaProTyProPhePr 597
932 AATGTTGATACCCAGACACAGCAGTGAATTTGTCCAGAGCTCCGATGGG 981
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
597 oProVal.....ProThrValThrThrSerSerAlaThrLeuSerThrV 612
982 GTGCAGCAGCCCGCAGCTGAGCCCGCAGTGCAGCAGCAGCAGCAGCAGCT 1031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
612 allleAlaThrValAlaSerSerPro.....AlaGlyTyThrSthr 625
1032 ACAGACAGCTGAGCTGCCAGATGCTGCTCCCGAGTCCAGCTGAGCAGC 1081
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 AlaSerProGlyProProProTyGlyGlyAlaGlyAlaProSer..... 640
1082 AGAGCAGCTCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
641 .....ProGlyAlaTyThrAlaThrAlaThrProProGlyTyGlyP 654
1122 .....GCAGTGCAGACCCCGCAGTGCATGCGC 1148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 roGlySerProProSerPheArgThrGlyThrProProGlyTyArgGly 670
1149 CCTCCCGCCCGCAGCTGCCG.....CAGCCCGCCAGCCAGCT 1189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
671 ThrSerProProAla.GlyProGlyThrPheGlySProGlySerProThyV 687
1190 CAGAGCCCACTCCAGCTGAGCTGCGCCCTGCCAGCCATCTCCAGTACG 1239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
687 alGlyProGlyProLeuProProAlaGlyProSerGlyLeuProSer... 702
1240 TTCCTGCCAGCCCTCCAGCGCAGCCCTCCAGAGCCAGTGCAGCGCGG 1289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 ...LeuProProProProAlaAlaProAlaSerGlyProProLeuSerAl 718
1290 GACC.....CCACAGAACTTCAGTGTGCC...TCAC 1318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
718 aThrGlnlLeuGlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 735
1319 CT...GGACCTTTAAACACAGCTGTGAACCCAGCTGTGCATGAGC... 1362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
735 roValProProAlaArgSerProSerProProProGlyValValPargVal 751

```

```

1363 CCAGCTGGCTCCAGCAG...GCTGAGAGAGCAGACTGACGACAG.. 1407
1407 ..... 1407
752 ProserhisAlaSerHisAlaArgPheasnLysHisLeuAspArg1 768
1407 ..... 1407
768 yPheasnSerCysAlaArgSerAspLeuArgPheValProLeuGluGlyS 785
1408 .....CTGAGCAGCTGCTGAGAGTATGACGAGCCCTGCGCGCAT. 1448
785 erLysLeuAlaLysLysArgAlaAspLeuValGluLysValArgArgL 801
1449 ...GATCAACAAGATGACAGACGAGACGAGAAAGAAAGAGCTGAGTA 1495
802 AlaGluGlnArgAlaArgGluGluGluGluGluGluGluGluGluGlu 818
1496 AGATGAAGAGCTGCTGAGCATTCGACAGACCCCTGACAGCGGTGCC 1545
818 gLulysGluArgLulysGluGluGluGluGluGluGluGluGluGlu...Ser 833
1546 CTGAAGACCTGCAAACTGAGATGCGCCCTGAGAACTCAAGAATGA 1595
834 ValLysLeuAlaGluGluGluGluGluGluGluGluGluGluGluGlu... 845
1596 CATGGCGTGC.....CCAGTCCCGCCAGCCGCGCGGTGCCAGC 1633
846 .....CysProSerLeuGluGluProValProHisArgProPro...PheG 859
1634 GACCAACAGCAGTATGACGAGCCGCTGATGATGCGCTGCTGCGCA 1683
859 LuProGluSerAlaValAlaThrValProProGluGluGluGluGluGlu 875
1684 ACATCGGTCACCTGCTTCAACCATTCCTGACCGCAGCATTCGTCGC 1732
875 rProAlaLeuArgThrLeuSerGluGluGluGluGluGluGluGluGlu 891
seq_name: SwissProt_39:LP61_E1MTE

seq_documentation_block:
ID LP61_E1MTE STANDARD; PRT: 255 AA.
AC P15714;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE ANTIGEN LPMC-61 (FRAGMENT).
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC NCBI_TaxID=5802;
RX MEDLINE=90348718; PubMed=2200963;
KO C., Smith C.K. II, McDonnell M.;
RT "Identification and characterization of a target antigen of a
monoclonal antibody directed against Eimeria tenella merozoites.";
Mol. Biochem. Parasitol. 41:53-64(1990).
CC -1- FUNCTION: UNKNOWN. THE GLN-RICH TANDEM REPEATS MAY BE IMPORTANT
FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN
IMPORTANT IMMUNOGEN.
CC -1- SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER
POLYPEPTIDES TO FORM THE 80 KDA ANTIGEN.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE
SPOROULATION OF THE OOCYSTS AND IN THE SPOOROZITES FOLLOWING
EXCystation.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: M30933; AAA29079.1;
DR PIR: A60637; A60637.
KW Antigen: Sporozoite; Repeat: Sporulation.
FT NON_TER 1
FT DOMAIN 1 18 210 12 X APPROXIMATE TANDEM REPEATS, GLN-
FT REPEAT 18 48 1.
FT REPEAT 49 57 2.
FT REPEAT 58 65 3.
FT REPEAT 66 78 4.
FT REPEAT 79 90 5.
FT REPEAT 91 103 6.
FT REPEAT 104 140 7.
FT REPEAT 141 152 8.
FT REPEAT 153 164 9.
FT REPEAT 165 172 10.
FT REPEAT 173 192 11.
FT REPEAT 193 210 12.
FT NON_TER 255
SQ SEQUENCE 255 AA; 31267 MW; 8C5E605FFFC2DB3 CRC64;

alignment_scores:
Quality: 310.00 Length: 190
Ratio: 2.719 Gaps: 7
Percent Similarity: 60.000 Percent Identity: 46.842

alignment_block:
US-09-668-119-1 x LP61_E1MTE
Align seg 1/1 to: LP61_E1MTE from: 1 to: 255
367 CCAGCAGCCAGCTGCTGAGCTCCAGCAGGTGGCGCTGACAGCAGCAGACA 416
27 ProGluProThrLeuAspArgGluGln.....GlnGlnGlnGln 40
417 ACAGCAGCAGGTTCCAG..... 432
40 sasnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 57
433 .....CAGCAGCAGCAGCGCGCGCTACAGCAGCAGCAGCAG 471
57 euTrpProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAla 73
472 CAGCAGCAGCAGCAGTTCAGGCTCAGCAGAGAGTGCCATGACAGCAGACT 521
74 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 88
522 CCAGCAGTAGTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG... 567
89 GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 105
568 ...CAGCAGCAGCAGTCTAAATTGATTCATCATCAAAATCAGCAGCAGATA 615
105 roGlnGlnGlnHis.....ArgArgGlnHisGlnGlnGlnGlnGlnGln 119
616 CAGCAGCAGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 665
120 MetAsnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 136
666 A...CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTT 712
136 nLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 153
713 TGAGAGCCAGCCAGCAATTCAGCAGCA.....CCGATGACAGCAG 753
153 rProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 169
754 CCAGCAGCTCCGCGCTCCAGAGGCTGCTGCCAGCAGCAGCTGACAGAGATGA 803
170 GlnGlnGlnGlnTrpSerAspGlnAsnGlnGlnGlnGlnGlnGlnGlnGln 186

```


481 CAGAGTTCAGAGCTCAGAGTCCATGCTCAGAGCAGAGTTCACAGCAGT 530
 |||||
 3 GlnGlnPheGlnAlaGlnGln*****AlaMetGlnGlnGlnPheGlnAlaVal 19
 |||||
 531 AGTGCAGCAGCAGCAGCAGCAGCTCCAGAGCAGCAGCAGCAGCAGCAGT 580
 |||||
 19 lValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 36
 |||||
 581 TAATTAATTCATCATCAAAATACAGACAGATACAGCAGCAGCAGCAGCAG 630
 |||||
 36 euilelylsleuHishlsGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 52
 |||||
 631 CAGCTGCAGGAATAGCAGCAGCTGCAGCTCCAAACAGCAGCAGCAGCAGCA 680
 |||||
 53 GlnleuGlnArgTyleAlaGlnleuGlnleuGlnleuGlnleuGlnleuGln 69
 |||||
 681 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCTTTGGAGGCCAGCCAGCA 730
 |||||
 69 nglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 86
 |||||
 731 TTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCCGCTCCAGGCTCTG 780
 |||||
 86 lGlnGlnGlnProPromeGlnGlnProGlnProPProPProGlnAlaLeu 102
 |||||
 781 CCCCAGCAGCTGCAGCAGATGCATCAACACAGCAGCAGCAGCAGCAGCAGC 830
 |||||
 103 ProGlnGlnleuGlnGlnMetHishlsThrGlnHishlsGlnProProPr 119
 |||||
 831 ACAGCCCCAGCAGCAGCTCCAGTGTGCTCAAAACCAACATCAACACTCCGC 880
 |||||
 119 OGlnProGlnGlnProProValAlaGlnAsnGlnProSerGlnleuProP 136
 |||||
 881 CACAGTGCAGACCCAGCTTTGGTGTGCACAGGCGCAAGCTCCCTCCGGA 930
 |||||
 136 roGlnSerGlnThrGlnProleuValSerGlnAlaGlnAlaLeuProGly 152
 |||||
 931 CAAATGTTGATATACCAACACCACTGAATTTGTCGAGCTCGATGATG 980
 |||||
 153 GlnMetleuTyrrThrGlnProProleuLysPheValArgAlaProMetVa 169
 |||||
 981 GGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1030
 |||||
 169 lValGlnGlnProProValGlnProGlnValGlnGlnGlnGlnAlaVal 186
 |||||
 1031 TACAGACAGCTGCAGCTGCAGATGGTGGCTCCGAGTCCAGCTGAGC 1080
 |||||
 186 alGlnThrAlaGlnAlaAlaGlnMetValAlaProGlyValGlnValSer 202
 |||||
 1081 CAGAGCAGCAGCTCCCATGCTGTCTCGCCGCTCACCGGCGCAGCAGGTGCA 1130
 |||||
 203 GlnSerSerleuProMetleuSerSerProSerProGlyGlnGlnGlnGln 219
 |||||
 1131 GACCCCGCAGTCGATGCCCCCTCCCGCCAGCCGTCGCCGAGCCGCGGCC 1180
 |||||
 219 nThpProGlnSerMetProProProProGlnProSerProGlnProLys 236
 |||||
 1181 AGCCAGCTCAGCAGCCCACTCAGCTGCAGCTGGCCGCCCCCAATCT 1230
 |||||
 236 lmpProSerSerGlnProAsnSerSerValSerSerGlyProAlaProSer 252
 |||||
 1231 CCCAGTAGCTCTGCGCCAGCCCTCCAGCCGAGCCCTCCAGAGCCAGT 1280
 |||||
 253 ProSerSerPheleuProSerProSerProGlnProSerGlnSerProVa 269
 |||||
 1281 GACGGCGGGGACCCCAAGAACTTCAGTGTCCCTCCACCTGGAGACCTTAA 1330
 |||||
 269 lThAlaArgThrProGlnAsnPheSerValProSerProGlyProLeuA 286
 |||||
 1331 ACACACCTGTAAACCCAGCTCTGTCATGAGCCAGCCAGCTGCTCCAGCAG 1380
 |||||
 286 snThProValAsnProSerSerValMetSerProAlaGlySerSerGln 302
 |||||
 1381 GCTGAGAGCAGCAGTACTGAGACAGCTGAAGCAGCTGTGGAAGTACAT 1430
 |||||

|||||
 303 AlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 319
 |||||
 1431 CGAGCCCCCTGGCCCGCATGATCAACACAGATGCACACAGACAGCAGCA 1480
 |||||
 319 eGlnProleuArgArgMetIleAsnLysIleAsnLysLysAsnGlnLysArgL 336
 |||||
 1481 AAAGAGCCTGAGTAGATGAAGAGCCTCTGAGC 1515
 |||||
 336 yslYAspLeuSerLysMetLysSerleuLeuAsp 347
 |||||
 seq_name: sp_invertebrate:09Y149
 seq_documentation_block:
 ID 09Y149 PRELIMINARY; PRT; 749 AA.
 AC 09Y149;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE BCDNA:GH03922 PROTEIN.
 GN BCDNA:GH03922 OR CG4184.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Piankovich C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Modarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,
 RA Palazolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster."
 Science 287:2185-2195(2000).
 [2]
 SEQUENCE FROM N.A.
 RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
 RA Agbayani A., Arcina T.T., Baxter E., Blazey R.G., Butenoff C.,
 RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,

Fri Mar 1 09:12:00 2002

us-09-668-119-1.rpt

Page 6

```
592 LeuAlaProSerAsnThrSerProProGlnGlnGlnGlnGlnGlnGln 608
564 GCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 613
608 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 615
614 TAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 663
615 aGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 628
664 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 713
629 ProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 645
714 GGAGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 741
645 nHISGlnValProGlnThrHisLeuHisGlnAlaAlaLeuSerGlnPro 662
742 .....CCGATGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 783
662 iStyValProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 675
784 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 807
676 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 692
808 .....ACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 829
692 nLeuValGlnThrHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 709
830 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 876
709 roGlnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 725
877 CCG.....CCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 905
726 ProLeuProIleThrHisThrMetProProGlnThrSerProVal 742
906 GTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 955
742 .....TGAATTTGTCGAGCTCCGATGTCGTCGTCGTCGTCGTCG 1005
743 .....ValThrSerProValLeuLeuGlnGlnGlnGlnGlnGln 756
1006 CAG.....GTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1049
757 MetProValValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 772
1050 CCAAGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1095
772 sProGlnValSerProAlaProProSerSerAsnThrThrThrPro 789
1096 ..ATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1143
789 LyIleAlaSerThrProThrAlaGlyValIleAlaSerAlaGly 805
1144 ATGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGC 1193
806 LysThrThrProAlaAlaProThrProThrSerAsnSerAlaThr 822
1194 GCCCACTCCCAAGCTGAGCTGTCGTCGTCGTCGTCGTCGTCGTCG 1243
822 oThGlyThrAlaAlaThrAlaGlyValAlaThrGlyThrThrPro 839
1244 TGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1293
839 aLysLysHisValLeuAsnProSerAlaLysProPheThrProAl 855
1294 GCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1343
856 Pro.....SerThrProAsnProSerAlaProHisThrProGln 869
```

```
1344 CCCAGCTCT..... 1353
869 rProValProMetThrAsnIleThrThrThrGlyGlnHisValPro 886
1354 .....GTCATGAGCCCA..... 1365
886 roAlaAlaAsnGlnProIleThrValMetGlnProGlnHisProPhe 902
1366 .....GCTGAGC..... 1371
903 ProGlnThrHisProGlnAlaGlyGlnProProAlaGlyLeuArgSer 919
1372 .....TCCACCCAGCCTGAGG 1387
919 nPyrProProMetAlaAlaSerGlnMetHisValSerAlaSerAla 936
1388 AGCAGCAGTACTGGACAAGCTGAGCAGCAGCTGCAAGTACATGAG 1436
936 hGlyGlnProLeuIleThrAlaGlyProIleProGlnPheIleGln 952
1437 ...CTGCGCCCGCATGATCAACAAGATGCAACAAGAACAGACAAA 1483
953 GlnHisAlaProHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 962
1484 AGGACCTGATGAATGAAGACCTTGTGACATTTCTGACAGACCCCT 1533
963 .....SerHisThrTyAlaProMetG 970
1534 AA.....GCGGTGTCCTGAGACCTTGAAGACCTTGAAGACCTGAG 1571
970 ImeTarValTyProAspGlnProGlnGlnGlnGlnGlnGlnGln 986
1572 CGCCCTGGAGAACT.....CAGATGACA 1597
987 ThrProGlnSerThrThrProSerProGlyGlnProHisGlnGln 1003
1598 TGGGTGTCCTGAGACCTTGTGACATTTCTGACAGACCCCTGAG 1619
1003 sProProProGlnProSerProAlaGlyGlyGlyProGlnProAla 1020
1620 ..CCCGCGTCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1664
1020 hProProThrGlnAlaAlaThrTyGlnLeuMetCysValHisPro 1035
1665 CTGATGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCT 1685
1036 ...GlnSerLeuLeuAlaAsn 1041

seq_name: sp_invertebrate:O90957

seq_documentation_block:
ID O90957 PRELIMINARY; PRT; 2123 AA.
AC O90957;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ADENYLYL CYCLASE.
GN ACPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Soderbom F., Anjard C., Iranfar N., Loomis W.F.;
RT "An adenylyl cyclase that functions during late development of
RL Dictyostelium."
CC -i- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSOR TRANSDUCTION SYSTEMS.
DR EMBL: AF153362; AAD50121.1; -
HSSP: P19754; 1AWK.
```

DR InterPro: IPR000410; Bcrl_sen
 DR InterPro: IPR001054; Guanylt_cyclase.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR001596; Pyrophosphatase.
 DR InterPro: IPR001789; Response_reg.
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00211; guanylate_cyc. 1.
 DR Pfam: PF02518; HATPase_c. 1.
 DR Pfam: PF00072; response_reg. 1.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR SMART: SM00044; CYC. 1.
 DR SMART: SM00387; HATPase_c. 1.
 DR SMART: SM00448; REC. 2.
 DR PROSITE: PS0125; GUANYLTATE_CYCLASES_2; 1.
 DR PROSITE: PS00387; PPASE; UNKNOWN_1.
 DR Phosphorylation; Sensory transduction.
 KW SEQUENCE 2123 AA; 243039 MW; 744247140BC342PA CRC64;

alignment_scores:

Quality: 456.50 Length: 318
 Ratio: 2.609 Gaps: 6
 Percent Similarity: 55.031 Percent Identity: 41.509

alignment_block:

US-09-668-119-1 x Q9U9S7 ..

Align seg 1/1 to: Q9U9S7 from: 1 to: 2123

379 CTGCGAGCTCCAGCAGGTGGCGCTGCAGCAGCAGCAGCAGCAGT 428
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 1826 LeuGlnLeuGlnIleGlnLeuGlnIleGlnLeuGlnIleGlnLeu 1842
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 429 CCAGCAGCAGCAGCAGCGCGGCTACAGCAGCAGCAGCAGCAGC 478
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1842 uGlnGlnGlnGlnGlnIleuGlnLeuGlnGlnGlnGlnGlnGln 1859
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 479 AGCAGCAGTTCAGGCTCAGCAGAGTGCATGCGCAGCAGCAGTTC 528
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1859 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1875
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 529 GTAGTGCAGCAGCAGCAGCAGTTCAGCAGCAGCAGCAGCAGCAG 578
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1876 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1892
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 579 TCTAATTGAATTCATCATCAAAATCAGCAACAGATACAGCAG... 625
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1892 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1909
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 626 AACAGCAGCTGCAGCAATAGCAGCAGTGCAGTCCCAACACAGCA 675
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1909 eGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGln 1925
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 676 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 725
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1926 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1942
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 726 ACCAATTGAGCAGCAGCAGTGCAGCAGCAGCAGTCCGCCCTCC 775
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1942 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1959
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 776 CTCTGCCCGCAGCAGTGCAGATGCATACACACACACACACAGCC 825
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1959 InLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1972
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 826 CCACACAGCCCGCAGCAGCTTCAGTTCAGAACCAACATCACA 875
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1973 SerProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1982
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 876 CCGCGCAGTGCAGCAGCAGCTTGGTGTACAGGCGCAGCAGCTCC 925
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1982 nProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1994

926 CTGAGCAAAATGTTGATACCAACACACACATGAATTTGTCGAGCTCCG 975
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1995TyrGlnHisGlnLeuGlnIleGlnIleGlnIleGlnIleGln 2005
 976 AAGGTGTGCAGCAGCCCCCAGTGCAGCCAGCAGGTGCAGCAGCAGCAG 1025
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2006 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2022
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1026 AGCAGTACAGACAGCTCAGGCTGCCAGATGGTGGCTCCGAGTCCAGG 1075
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2022 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2039
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1076 TCAGCCAGAGC.....AGCTCCCATC 1098
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2039 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2055
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1099 CTCTCTCCGCGTCCAGCGGCGCAGCAGGTGCAGACCCCGCAGTCCAG 1148
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2056 PheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2072
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1149 CCTCCCGCCCGCAGCGCCGCGCAGCCGCGCAGCTCAGACCCCA 1198
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2072 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2089
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1199 ACTCCAGCTCAGCTCTGCGCTGCCCATCTCCAGTCTCCAGTCTGCC 1248
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2089 LegInLysSerGln..... 2094
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1249 AGCCCTCACCAGCCCTCCAGAGCCGAGTACGCGCGGAGCCGAC 1298
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2095 HisProHisSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2111
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1299 GAGC 1302
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2111 nAsn 2112
 : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 seq_name: sp_human:014687
 seq_documentation_block:
 ID 014687 PRELIMINARY; PRT: 4957 AA.
 AC 014687;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ALR.
 GN ALR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97388474; PubMed=9247308;
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
 RA Rallepalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
 RA Canaan E.;
 RT "structure and expression pattern of human ALR, a novel gene with
 RT strong homology to ARL-1 involved in acute leukemia and to Drosophila
 RT trithorax.";
 RL Oncogene 15:549-560(1997).
 RL EMBL: AF010404; AAC51735.1; -;
 DR InterPro: IPR000910; HMG_12_box.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR001822; Recombinase.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR002965; P-rich_extensn.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR003888; FYrich_N.
 DR InterPro: IPR003889; FYrich_C.
 DR Pfam: PF00628; PHD; 3.
 DR Pfam: PF00856; SET; 1.
 DR PRINTS: PR01217; PRICHEXTENSIN.
 DR PROSITE: PS00396; RECOMBINASES_2; UNKNOWN_1.


```

3738 1yAlaLeuGlyProThrLeuLeuLeuThrGlyLysGluGlnAsnThrVal 3754
750 .....
3755 AspProAlaValSerSerGluAlaThrGlnGlyProSerThrHisGlnG 3771
750 .....
3771 yGlyProLeuAlaIleGlyThrThrProGluSerMetAlaThrGluProG 3788
750 .....
3788 1yGluValLysProSerLeuSerGlyAspSerGlnLeuLeuValGln 3804
751 ...CAGCCAGACCTCCGCCCTCCAG.....GC 776
3805 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 3821
777 TCTGCCCCAGACGCTCCAGCAG.....ATGCAT..... 804
3821 GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3838
804 .....
3838 1yGlySerHisGlyGlnLeuGlySerSerSerSerSerSerSerSerSer 3854
804 .....
3854 ValProHisLeuLeuAlaGlnProSerValSerLeuGlnLysGlnProG 3871
805 ...CAGACAGACGACCCAGCCGACCCAGACCCAGACCCAGCCTCCAG 850
3871 ySerMetThrGlnAsnLeuLeuGlyProGlnGlnProMetLeuGlnArg 3888
851 TTGGTCAGACCAACACCACTCCGCGACGACGACGACGACGACGACCT 900
3888 rometGlnAsnAsnThrGlyProGlnProPro.....LysProGlyPro 3902
901 TTGGTCAGACGCGGCAAGCTCCGTCGCAATATGTTATATACCCAGC 950
3903 ValLeuGlnSerGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3917
951 ACCACTGAATTTGCTGAGCTCCGATG.....GTGGTCAGACGACCC 994
3917 oThrValGlyGlnLeuArgAlaGlnLeuGlnGlnGlnGlnGlnGlnGln 3934
995 CA.....GT3CAGCCCGAGGTGACGACGACG..... 1020
3934 roGlnLeuArgHisLeuSerProGlnGlnGlnGlnGlnGlnGlnGlnGln 3950
1021 ...CAGACAGCAGTACAGACAGCTCAGCTGACCCAGATGGTG..... 1059
3951 LeuMetGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3967
1060 .....GCTCCCGGATTCAGGTCAGC.....CAGACAGACCTC..... 1092
3967 oTyGlnGlnProGlyThrGlnThrSerProLeuGlnGlnGlnGlnGln 3984
1093 ...CCCATGCTGCTCTGCGCTCCAGCCGACGACGACGACGACCCG 1137
3984 yGlnProGlnLeuGlyGly...PheProGlyProGlnThrGlyProLeu 3999
1138 CAGTCAGTGGCCCTCCGCCCGACGCTCCGACCCGACCCGACCCGAC 1187
4000 GlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4016
1188 CTCACAGCCCACTCCAGCTCAGCTGCG..... 1218
4016 a.....ProProGlyAlaLeuSerThrGlyProValLeuGlyProVal 4031
1219 ..CCTGCCCATCTCCATAGTCTCTGCCGACCC...CTCAGCCGACG 1263

```

```

4031 IsProThrProProProSerSer..ProGlnGlnProLysArgProSerG 4047
1264 CCTCCAGACGACCCAGTGAAGCGCGACCCAGACCACTGATGTCGCC 1313
4047 InLeuProSerProSerSerGlnLeu...ProThrGlnAlaGlnLeuPro 4062
1314 CTCACCTGACCTTTAAGACACACCTGTGAACCCAGCTGTCACGAGCC 1363
4063 Pro.....ThrHisProGlyThrProLysProGln.....GlyPr 4074
1364 CAGCTGCTCCAGCTGACGCTGACGACGACGACGACGACGACGACG 1413
4074 oThr...LeuGlnProProGlyArgValSerProAlaAlaGlnLeu 4090
1414 CAGCTGTCAGATTA...CATCGAGCCCTGCGCGCAGCATGATCAGAC 1460
4090 euAlaAspThrLeuPheSerLysGlyGlyProThrPaspProProasp 4106
1461 CGACAAAGAAACAGACAGAAAAAGACCTGAGTATGATGAGACGCTTC 1510
4107 AsnLeuAlaGlnThrGlnLys.....ProGlnGlnSerSerLeuValPr 4121
1511 TGACATCTCTGACAGACCCCTCGAAGCGGTGTCCTCCGAA 1550
4121 oGlyHisLeuAspGlnValAsnGlyGlnValValProGlu 4134
seq_name: sp_invertebrate:045322

```

```

seq_documentation_block:
ID 045322 PRELIMINARY; PRT; 796 AA.
AC 045322:
DT 01-JUN-1998 (Tremblere, 06, Created)
DT 01-JUN-1998 (Tremblere, 06, Last sequence update)
DE 01-JAN-1999 (Tremblere, 09, Last annotation update)
DE DYS 5.5 PROTEIN.
GN DYS 5.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Leonard N.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RX [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Lettelle P.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sultson J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan R., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RL Nature 368:32-38(1994).
DR EMBL, Z96047, CAB09418.1, -.
SQ SEQUENCE 796 AA; 86783 MW; 4FDFCB8B5461059 CRC64;

```

```

alignment_scores:
Quality: 447.00 Length: 509
Ratio: 1.781 Gaps: 24
Percent Similarity: 49.312 Percent Identity: 32.024

```

```

alignment_block:
US-09-668-119-1 x 045322 ..
Align seg 1/1 to: 045322 from: 1 to: 796

```



```

1071 GACTCGGGGAGGACCATCTGGGAGGCTGCTGCTGCTGCTGCT 1022
43 .....CysLysProValC 47
1021 GCTGCTGC.....TGACCTGGGCTCCTGAGGCTGCTGACAC 981
|||||CysCysLysProValCys.....CysCysVal 56
47 yscyscysLysProValCys.....CysCysVal 56
980 ACATCGAGCTGGG/CAAAATTGATGCTGGTGGTATACAAATTGG 931
57 ProVal..... 58
930 TCAGAGGAGAGCTTGCAGCTGACACCAAGGCTGGGCTGCGACTG 881
59 .....CysSerCysSerCysGly.....GlyCysG 68
880 GCGGAGATTGTGATGCTTGGTTGAGCACTGAGGCTGCTGGGCTGT 831
|||||GlyCysGlySerCys.....GlyCysGlySerCys 77
68 LysSerCys.....GlyCysGlySerCys 77
830 GGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781
78 GlySer.....SerCys..... 81
780 CAGAGCCGGAGGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCT 737
82 .....GlyCysGlySerCysCys..... 89
736 GCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 696
90 .....LysProValCysCysValProVal 98
695 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 669
99 CysSerCysSerCysGlyGlyCysLysProCysCysCysGlnSer 115
668 .TGTGTTGAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 620
|||||CysCysLysProCys.....CysSerCysGlySerCysCys 131
115 rCysCysLysProCys.....CysSerCysGlySerCysCys 131
619 GCTATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 570
131 InSerCysCys.....LysProCysCysCys 140
569 .....TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
141 GlnSerCysCysLysProCysCysGlnSerCysCysLysPr 157
539 .TGTGCTGACTACT..GCTGGAACCTGCTGCTGCTGCTGCTGCTG 494
|||||CysCysLysProCys.....CysSerCysGlySerCysCys 172
157 oCysCysSerSerCysGlySerCysCysGlnSerCys..... 172
493 CCTGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 459
173 .....CysLysProCysCysGlnSerCysCysLysProCys 186
458 TCGTCTGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
|||||TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
187 CysCysGlnSerCysCysLysProCysCysGlnSerCysCys 203
417 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 374
203 sLysProCysCysLysProCysCysLysProCysCysSerSer 220
373 TCTGTGGA 366
220 LysCysGly 222

```

seq_name: sp.invertebrate.Q9VUB6

seq_documentation_block:

ID Q9VUB6 PRELIMINARY: PRT: 1798 AA.

```

AC Q9VUB6;
DT 01-MAY-2000 (TRENDEL 13, Created)
DT 01-MAY-2000 (TRENDEL 13, Last sequence update)
DT 01-JUN-2001 (TRENDEL 17, Last annotation update)
DE CG8797 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrita J.F., Agbayant A., An H.-J., Andrews Plennoch C., Baldwin D.,
RA Ballew K.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova K.C., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Butris J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durdin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegyan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulyov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pittman K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Stupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zakeri J.S., Zhao W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2193(2000).
DR EMBL: AE003556; AAF4972.1; -.
DR FLYBase: FBgn0036399; CG8797.
DR InterPro: IPR001357; BRCT.
DR Pfam: PF00533; BRCT; 4.
DR SMART: SM00292; BRCT; 4.
DR PROSITE: PS50172; BRCT; 2.
SQ SEQUENCE 1798 AA; 205695 MW; 64E4963B1181B69 CRC64;

```

alignment_scores:

Quality: 427.50 Length: 620
Ratio: 1.469 Gaps: 26
Percent Similarity: 46.935 Percent Identity: 29.194

alignment_block:

US-09-668-119-1 x Q9VUB6 ..

Align seg 1/1 to: Q9VUB6 from: 1 to: 1798

283 CAGCCATGCTCTCTCAGGAGGAGCCGCTCTGAGGAGCTGCGGAGATGCC 332

```

336 GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnHisThrProSe 352
333 CCCACAGCATGGCTGCTGCTAGCGACATCCACAGACC..... 375
352 rProArgLInSerProLeuGlnGlnProThrThrProThrLeuGln 369
376 .....CAGCTGCAGTCCAGAGGTGGCG 399
369 InGlnProAsnGlnGlnAsnAlaGlnGlnGlnGlnGlnGln 385
400 CTGACAGACAGACAGACAGACAGACTTCCAGACAGACAGAGCGGC 449
386 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnVa 402
450 GCTACAGACAGACAGAG.....CAGCAGACAGCAA..... 477
402 lLeuThrGlnGlnProGlnProGlnGlnGlnGlnValIleThrG 419
478 .....CAGCAGACAGTTCCAG 492
419 lnaArgHisValIleAsnThrSerThrAlaGlnGlnGlnGlnIle 435
493 GCTACAGACAGATGGCATGACAGACAGATTCCAGACAGTAGTCAG 542
436 SerHisMetSerLeuAlaLeuGlnGlnGlnGlnGlnGlnGlnGln 452
543 GCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCATCTAATTAAATG 552
452 nGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnIleThrGlnVal 469
593 ATCAT.....CAAAATCAGACAGATATACAGCAGCAGCAGCA 630
469 lInGlnLeuProProAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 485
631 ...CAGCTGCAGCAATAGCAGACGCTGCAGCTCAACACAGCAA.... 672
486 ValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnI 502
672 ..... 672
502 eAlaLeuGlnAlaGlnGlnGlnValArgIleGlnGlnThrIleGln 519
673 .....CAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAG 705
519 rGProProThrAsnValGlnGlnGlnGlnGlnValProGlnGlnGln 535
706 CAGGCTTTGGAGCCAGCCAGCCAAATTCAGACAGCAGCAGCAGC 755
536 GlnGlnIleIleGlnGlnPheProGlnGlnGlnGlnGlnGlnGln 552
756 ACAGCTCCGCTCCAGAGCT...CTGCCAGCAGCTGCAGAGCAATGC 802
552 oGlnGlnGlnIleSerGlnValGlnGlnGlnGlnGlnGlnGlnGln 569
803 ATCACA.....CAGCAG...CAGCAGCCGCCAGCAGCAG 834
569 lInyThrPheIleIleSerGlnGlnHisPheAsnGlnLeuSerVal 585
835 CCCAGCAGCAGCTCCAGTGTCTCAGAACACACATCAGAA..... 873
586 GlnGlnGlnGlnIleLeuAlaGlnAsnProGlnSerGlnAsnVal 602
873 ..... 873
602 lValAsnGlnThrGlnGlnGlnLeuValGlnGlnGlnGlnGlnGln 619
874 .....CTCCGCGCAGATGCCAGACCCAGCAGCTTTGGTGTGCAG 912
619 lInGlnGlnIleIleProGlnGlnGlnGlnGlnGlnGlnGlnGln 633
913 GCGCAGAGCTCTCCCTGCAGCAAAATGTTGTAGCCAA..... 948

```

```

634 MetGlnGlnGlnProGlnGlnIleValValAsnGlnGlnIleLeuSer 650
949 ...CCACACTGAATTTGTCGAGACTCCGATGGTGTGCAG...CAGC 991
650 pProGlnArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 667
992 CCCAGTGCAGCCCGCAGTGCAGCAGCAGCAGCAGCAGCAGTACAG 1041
667 lInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 683
1042 CAGGCTGCCAGATGGTGGCTCCCGAGTCCAG.....GTCCAGCA 1082
684 GlnGlnGlnGlnLeuValGlnGlnGlnGlnGlnGlnGlnIleLeu 700
1083 GAGCAGCCTCCCAATG.....CTGCTCTGCG 1108
700 nGlnGlnThrProGlnGlnGlnGlnIleLeuGlnGlnGlnGlnSer 717
1109 CGTCACCCGCGCCAGCAGCTG.....CAGACCCCG 1137
717 sPAsnGlnGlnGlnGlnIleMetGlnLeuIleThrGlnGlnGlnPro 733
1138 CAGTGCATGCCCTCCCTCC..... 1155
734 GlnArgThrProProProGlnGlnGlnGlnGlnGlnGlnGlnGln 750
1156 .....CCGACGCTCCCGCAGCCGCGCAG..... 1182
750 nGlnGlnGlnGlnSerProGlnHisHisProGlnMetGlnGlnHis 767
1183 ....CCGAGCTCAGACCCCAATCCAGCTCAGCTGGCTCCGCTCC 1227
767 rPserProGlnSerProAlaGlnGlnGlnMetAlaSerThrProGly 783
1228 TCTCCAGTAGCTTC.....CTGCC.....AG 1250
784 ThrProThrSerValGlnMetGlnSerProLeuProGlnGlnPro 800
1251 CCCCTCAGCCGAGCCCTCCAGAGCCAGTGCAGGCGGAGCCCAAGA 1300
800 rProGlnProGlnProGlnGlnGlnPheValProArgGlyLeuArg 817
1301 ACTTCAGTGTCCG.....TCAGCTGCAGCTTTAAACAGCACT 1338
817 lInThrSerThrProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 833
1339 GTGAACCCAGCTCTCATAGCCAGCCAGCTGCTCCAGCAG...GTGA 1385
834 GlnGlnGlnGlnIleValLeuProProProGlnGlnGlnGlnGln 850
1386 GGAGCAGCAGTACCTGGAACAAGCTGAAGCAGCTGTCAATATCGAG 1435
850 nGlnGlnGlnHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 866
1436 CCCTGCGCCGATGATCAACAAGATCGACAAGAGAGCAAGAAAAAG 1485
866 rGlnGlnGlnGlnIleValValAspGlnGlnGlnGlnGlnGlnGln 882
1486 GACCTGAATGAATGAAGAGCTTCTGAGCATTTCTGACAGACCCCTGA 1535
883 SerLeuAspAlaHisGlyArg...AlaGlnGlnGlnGlnGlnGln 898
1536 GCGGTGTCCCTGAAAGACCTTGCAGAAAGTGAATGAGCTGAGAAAC 1585
898 nGlnAsnGlnProArgValMetLeuArgGlnGlnGlnGlnGlnGln 913
1586 TCAAGAAATGACATGGTGGCTCCCTCCAGCCAGCCGCTGCGCAGCA 1635
914 .....Pro.ArgProGlnAlaProGlnGlnGlnGlnGlnGln 923
1636 CCAACAGCAGTACCTATGCGCAGCCGCTCTGAGTGGCTGGGCAAC 1685
923 lAlaThrArgProAlaGlnProValProProGlnGlnGlnGlnGln 940

```


[illegible]

GN CG7177 Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Phlebotominae; Diptera; Tephritidae; Bactroidea; Muscomorpha;
OC Ephyridae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BKERY;
RX MEDLINE-20196060; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballen R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
RA Beeson K.Y., Bestou P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan D.A., Boutz J., Brockstein P., Brottier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cayley S., Dahlke C., Davernport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Deyanov I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glaser C., Gabriellian A.E., Gary N.S., Gelbart M.M., Glasser K.,
RA Glöckner C., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jalali M., Kalush F., Kaplen G.B., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Lin X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J.C., Mosherfi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spiter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svyrkas R., Tector C., Turner R., Venler E., Wang A.H., Wang X.,
RA Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavertil J., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003594 ARS1744.1,
DR FLYBase: FBgn0037098; CG117,
DR InterPro: IPR000719; Euk_Pkinase.
DR Pfam: PF00069; Pkinase_2
DR PROSITE: PSS0001; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.

SEQUENCE 2285 AA; 252559 MW; OE94A907CA01C3BA CRC64;

```

alignment_scores:      Quality: 423.00      Length: 604
                        Ratio: 1.488      Gaps: 33
Percent Similarity: 47.185      Percent Identity: 29.967

alignment_block:
US-09-668-119-1 x Q9VP17 ..

Align seg 1/1 to: Q9VP17 from: 1 to: 2285

286 CCAATGTCCTCTCATGGGACGCCCTCTCTGGACATCGGGATGCGCCC 335
|||||:.....:||||| |||
945 PRGSLGVALGVALGVALGSLGSLPRGSLTHRYALLAUPROPLNPR 961
|||:.....:||||| |||

336 TCACAGCATGCTGTCTGTCTACGGCACAATCCACACCCAGCTGCAGC 385
|||||:.....:||||| |||

```

```

961 OHISGLuIn.....GlnProGlnGlnGlnGln 972
386 TCCAGCAGGTGGCGTCGACAG..... 408
972 roLeuGlnGlnGlnGlnLeuMetHisThrAsnValGlnAlaPro 988
409 .....CAGCAGCAGCAGCAGCAGTTCACAGCAGCAGCAGCAGCGCGC 452
989 AspLeuThrGlnGlnGlnGlnMetAlaGlnGlnAlaGlnGlnIlePro 1005
453 ACAGCAGCAGCAGCAGCAGCAGAA.....CAGCAGCAGT 487
1005 eGlnGlnGlnGlnGlnGlnProGlnAlaValAlaAsnMetGlnGlnAla 1022
488 TCCAGGCTCAGCAGAGTCCAGCAGCAGTTCACAGCAGTTCAGAGCAG 537
1022 yrlAlaMetGlnGlnAlaGlyGlnGlnGlnGlnLeuSerGlnProLeuGln 1038
538 CAGCAGCAGCAGCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGTAA 587
1039 lLeGlnGlnGlnIleLeuGlnGlnGln.....ValAl 1050
588 ATTGCATCATCAAAATCAGCAGCAGTATCAGCAGCAG.....CAACAGC 631
1050 aValSerHisGln...GlnGlnIleMetGlnGlnGlnLeuAlaGlnHisG 1066
632 AGCTGCAGGAATATGACAGCTCAGCTCCAAACAGCAGCAGCAGCAG 681
1066 lInLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1082
682 CAGCAGCAGCAGCAGCAGCAGCAGCAGGCTTGGAGCCACAGCAGCAAT 731
1083 GlnGlnGlnIleGlnGlnGlnGln.....Le 1091
732 TCAGCAGCAGCAGATGACAGCAGCAG.....CCTCCGCCCTCCAGG 775
1091 uGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAla 1108
776 CTCTGCCCCAGCAGCTGACAGCAG.....ATG 801
1108 lAeMetProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1124
802 CATCAGCAGCAGCAGCAGCAGCAG..... 825
1125 AlaProHisGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1141
826 .CCAGCAGCAGCAGCAGCAGCTCCAGT..... 852
1141 lProProThrSerValAlaProProIleGlnHisThrTyraGlnGlnGln 1158
853 .....GCTCAGAACCA..... 864
1158 lYglYglValThrLeuSerAspAlaGlnGlnGlnGlnHisProGlyPhe 1174
865 .....CCATCACAACCTCCGCCA.....CAGTCCAGACCCAG.. 897
1175 SerAlaValProGlnGlnAlaAlaProPheIleGlnGlnProThrGlnGln 1191
898 .....CCTTGGTGTCAAGCGCAGCAGCTCCCTG 928
1191 nProIleGlnLeuSerMetProLeuGlnGlnGln..... 1204
929 GACAAATGTTGTATACCAACCA..... 951
1205 ..GlnLeuLeuHisSerGlnProAlaGlnGlnGlnGlnAlaMetSerGln 1220
952 .....CCACTG...AAATTTGTCGAGCTCCGAGTGGTGTCA 986
1221 GlnGlnGlnGlnGlnProLeuValGlnGlnGlnGlnGlnGlnGlnGln 1237
987 GCAGCCCCAGTGCAGCCCCAGTGCAGCAGCAGCAGCAGCAGCAGCAG 1036
1237 nGlnProProLeu.....ValGlnGlnGlnGlnGlnGlnGlnGln 1251

```

```

1037 CAGCTCAGGCTCCCAAGATGTGGCTCCCGAGATCCAGTCAAGCCAGC 1086
1251 lSgGln.....ProSerValGlnHisGlnGln... 1260
1087 AGCCTCCCATGTGTGCTCCCGCGTACCGGCGCAGGTCGACAGACCC 1156
1261 .....ProLeuValGlnGlnProGln.....GlnGlnGlnProGlnPro 1273
1137 GCAGTCGATGCCCTCCCTCCCTCC.....CAGC 1162
1273 oGlnAsnGlnGlnProGlnProGlnGlnGlnGlnGlnGlnGlnGlnGln 1290
1163 GCTCCCGCCAGCGCCGCGCAGCCCACTCAGACCCCACTCCAGCTCAGC 1212
1290 lGlnProProGlnGlnGlnGlnProAlaValGlnGlnIleSerGlnIleSer 1306
1213 TCTGCGCCTGCGCCATCTCCAGTGTCTCGGCCAGCCCTCAGCGCA 1262
1307 Ser.....GlnValProValGlnGln 1313
1263 GCCCTCCAGACCCAGTACGCGCGGAGCCCAACAGAACTTCAGTGTCC 1312
1313 nGlnAsnLeuGlnProIleGlnValAlaAsnLysAspAlaAsnValAlaThr 1330
1313 CCTACCTGGACCTTTAAACACA.....CCTGTGAACCCCACTCT 1353
1330 sPAlaMetSer...LeuAsnSerAlaHisGlyAlaLeuGlnProAlaPro 1345
1354 GTCATGAGCCCGCAGCTCCAGCAGCAGCTGAGGACGACGAG..... 1395
1346 LysThrGlnProGlnAsnSerAlaAspAlaGlnGlnGlnGlnGlnGln 1362
1395 ..... 1395
1362 rGlyThrGlyThrArgSerGlnLysProArgArgSerAsnArgSerGly 1379
1396 .....TACCTGGCAACACTCAACAGC 1416
1379 sNGLuArgIleProLysLeuSerValThrSerValAspGlnGlySerVal 1395
1417 CTGTGGAATGATCATCGAG...CCCTGGCGCGCATGATCAACAGATGCA 1463
1396 lLeAsnGlnHisMetGlnAsnLysLeuLysThrIleThrPheLysPheAs 1412
1464 C.....AAGACGACAGCAGAAAAAAGACCTGATGATG.....A 1501
1412 rIleGlyAspArgAspGlnThrAlaSerAspIleThrLysMetPheGln 1429
1502 AGAGCCTTCTGACATCTGTACAGACCCCTGGAAGCGGTGTCCCTGAG 1551
1429 roThrIleHisGlyValGlnThrLeuProSerGlyGlyGlyAlaGln 1445
1552 ACCTTGCAAAAGTGTGAGATCGCCCTGGAGAACTCAAGATGACATGCG 1601
1446 Gln...SerAsnGlnAsnLeuThrLeuGln.....AlaArg 1456
1602 GTGCCCACTCCCGCCAGCGCCCGCGGTGCGACCGACA.....AAGACA 1645
1456 gSerHisLeuSerAsnIleProAsnAlaLysGlnProGlnLeuAsnVal 1473
1646 GTACCTATGCGCAGCGGCTCGATGCGCTGTGCGCAACATCGCGTAC 1695
1473 eThr.....ProProThrThrThrSerThrMet 1482
1696 CTGTCTTCAACC 1707
1483 SerSerSerSer 1486

```

```

seq_name: sp_invertebrate:Q9W3G1
seq_documentation_block:
ID Q9W3G1 PRELIMINARY; PRT; 926 AA.

```

AC Q9W3G1;
 DT 01-MAY-2000 (TREMBLER1.13, Created)
 DT 01-MAY-2000 (TREMBLER1.13, Last sequence update)
 DT 01-JUN-2001 (TREMBLER1.17, Last annotation update)
 DE CG10555 PROTEIN.
 GN CG10555.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan N.R., Bouck J., Brokstein P., Center A., Chandra I.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Dubin K.J., Evans G.L., Garg N.S., Gelbart W.M., Glasser K.,
 RA Fodor C., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Glodok A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam G.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro F., Lei Y., Leitesky A.A., Li J., Li Z., Liang X., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klanos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtk R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT Science 287:2185-2193(2000).
 RL "The genome sequence of Drosophila melanogaster."
 DR EMBL: AE003443; AAF45366.1;
 DR FLYbase: FBgn0030034; CG10555.
 DR InterPro: IPR002965; Pfich.extensn.
 DR PRINTS: PR01217; PRICHTEXTENS.
 SQ SEQUENCE 926 AA; 33004 MW; 6FD5B9F77C36C006 CRC64;

alignment_scores:
 Quality: 419.00 Length: 734
 Ratio: 1.350 Gaps: 41
 Percent Similarity: 41.962 Percent Identity: 28.883

alignment_block:

US-09-668-119-1 x Q9W3G1 ..

Align seg 1/1 to: Q9W3G1 from: 1 to: 926

148 GCTTCCTGATGATGCTATGATGACATGACAGGCTGACGCGGAC 197
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 189 AladInEnuLaasPrcAlaMeLaasNlleSerGlnlleLeuProProp 205

198 T.....GTCGGGAGCCGCTGGAATTCGATGCGC 229
 205 ONHSLLEuGlnlThGlnlMetGlnlGlnlGlnlThPro 222
 230 CTCGGGGCCCGGACAGCTCTGGGGCGG..... 258
 222 roThGlnlProNlsglMetLeuGlylProProGlnlGlnlGlnl 238
 259 ATGGGATGCTTGGGTCATGGAGCAGCAGCTCTCTCTGAGGAGCC 308
 239 ProGlnlGlnlProValProGlnlPro.....GlyGlnl 251
 309 GCTTCCTGGAGCTCGGGATGGCCCTGACAGATGCTGCTGCTCTA 358
 251 yProPro...GlnMetGlnMetGlnlNlsgl..... 261
 359 CGGAGATCTCCAG...ACCAGCTGAGCTCCAGAGTGGGGCTGAG 405
 262 ..GlyAspProGlnlGlyProProValGlnMetProProGlyGlnl 277
 406 CAGCAGCAGCA.....CAGCAGCAGTT 428
 278 GlnGlnProGlnlProNlsglProGlyLeuProProGlyAlaGlnl 254
 429 CCAGCAGCAGCAGCAGCGGGCTAGCAGCAGCAGCAGCAGCAGC 478
 294 rGlnlGlnlGlnlGlnl.....GlnlGlnlGlnlGlnlGlnl 309
 479 AGCAGCAGTTCTCAGCTCAGCAGATGCTGAGCAGCAGTTCAGCA 528
 309 GlnGlnGlnlGlnlGlnlGlnlAlaAlaAlaAlaAlaAlaAla 325
 529 GTAGTGCAGCAGCAGCAGCT..... 551
 325 lAlaAlaAlaAlaValAlaAlaGlnlGlnlGlnlGlnlProGlnlVal 341
 552CCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 592
 342 GlnAlaGlnlProGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnl 358
 593 ATCATCAAAATCAG.....CAGCAGATCAGCAGCAGCAG 624
 358 snAlaGlnlGlnlGlnlGlnlProGlyAlaGlnlGlnlGlnlGlnl 374
 625 CAACAGCAGTTCAGCAGATGACAGCAGCTGCTCCAGCAGCAGCA 674
 375 GlnlGlyProValGlnlSerValIleAsnProAsnAlaAlaProAsn 391
 675 G.....CAGCAGCAGCAGCAGCAGC 694
 391 nArProAsnAsnGlnlProLeuSerGlnlProGlnlAsnProGlnl 408
 695 AGCAGCAGCAGCAGCTTGGAGGCCGCCAGCAGCAGCAGCAGCAG 744
 408 GlnGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnl 424
 745 ATGAGCAG.....CCAGAGCT..... 762
 425 GlnGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnl 441
 763CCGCCCTCCAGGCT.....CTGCCAGCAGCAGCAG 793
 441 yProGlnlValProProProGlnlSerProGlyAlaGlnlGlnlGlnl 458
 794 AGCAGATGATCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 843
 458 GlnGlnlGlnlNlserIstyrProGlyTyrProGlnlGlnl..... 473
 844 CCTCCAGTTCAGAACCAACATCAGCACTCCGCCAGCAGTGCAG 893
 474ThGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnlGlnl 485

```

894 CCAGCCTTTGGTGTACAGAGCGCAAGCTCTCCCTGGACAATGTGTGATA 943
      |||
486 .....ProProThrlnglYtyrG 492
944 CCCAACCCACGCTGAATTTGTCCGAGCTCCGATGGTGGACAGACCC 993
      |||||
492 lYProProPro.....Pro..Gly.....Pro 499
994 CCAGTGCAGCCCGCAGTGA.....GCAGAGCAGACAGACAGACA 1034
      |||||
499 roAsnAlaAlaGlnGlnGlylYtyrHisHisGlyProAlaGlyAlaAlaThr 515
1035 GACAGCTCAGGC.....TGCCAGAGTGGTCCCGAGT..... 1070
      |||||
516 G1yAlaSerGlnHisGlyTyrglnProAsnAlaGlyAlaGlylnglYpr 532
      |||||
1071 ....CCAGTGCAGCAGACAGCTCC.....CATGCTGTCTGGCCG 1110
      |||||
532 oProProGlyAlaTyrrProProProProGlySerGlnGlnVal.ProPro 548
1111 TCACCGGCGCAGCAGGTGCAGACCCCGCAGTCGATGCCCTCCGCCCA 1160
      |||||
549 ValProGlylngln..GlnProProProGlyProProProProGly 564
1161 GCGG.....TGCCGCGAGCCGCGCAGCCAGCTCAC 1192
      |||||
564 nProProThrlnglYlYlGlnGlnProProProGlyProProGlnSerG 581
1193 AG.....CCCACTCCAGCTCAGCTGTGCGCT..... 1221
      |||
581 lntYrglYProProProGlnAsnSerAlaGlylYlYProProPromet 597
1221 ..... 1221
598 G1yTyrrAlaGlyTyrrProProAsnProGlylntYrglYlAlaGlyAl 614
1222 .....GCCCATCTCCAGCTTCTCTGCCAGCCCTCAGCCGACG 1264
      |||||
614 aglYglYlYProProProSerGlyTyrrProProProProThrs 631
1265 CCTGC...CAGAGCCGAGTGAGGCG..... 1287
      |||
631 erSerAlaGlnSerProTyrglnAlaTyrglnGlnGlnGlnGln 647
1287 ..... 1287
648 AlaAlaAlaGlylYlYlAlaGlyAlaProProGlySerSerTyrrProG 664
1288 .CGGACCCACAGACTTCAGTGTCCTCAGCTCAGCTGAGCT..... 1326
      |||
664 yglYProProThrsGlyAlaAlaProProProProGlyAlaAlaT 681
1326 ..... 1326
681 yrSerThrThrrAlaProSerGlnThrProProProGlnGlylYlYgly 697
1327 .....TTAAACACACTGTGAACCCAGCTCTGCATGAGC... 1362
      |||||
698 AlaGlylYlYlYAsnAsnAsnProAsnGlyProAsnAlaGlnGlnSerTh 714
1363 .....CCAG 1366
714 rProProProGlnGlylYlYlAlaGlylYlYlAlaGlyProSerGlyProG 731
1367 CTGGCTCCAGCCAGGCTGA..... 1385
731 lYglYAlaGlylnglnlntYrrAlaGlyProProGlnGlnGlnProGlnGln 747
1386 .....GAGCAGCAGTACTGGACAAGTGAAGCAGCTGTGCA 1423
      |||||
748 GlnGlnProProGlylYlYlAlaSerGlylYlAlaAlaProLeuProThrl 764
1424 AGTACT.....CGAG 1434

```

```

764 nValGlnProThrlYrrSerThrProGlylYAsnTyrrAsnGlnGlnProGlyAl 781
      |||||
1435 CCCCTGGCGCGCATGATCAACAAGATCGACAAGAAAGAAAGAAAAA 1484
      |||||
781 lAProProProProAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 797
1485 GAGCTGAGTAAAGATGAAGAGCTTGTGACATTTGTGACAGACCCCTGCA 1534
      |||||
798 GlnGlnGlnGlnThrProProSerAlaGlylYlYlAlaGlylYlYl 814
1535 AGCGTGTCCCTGAAGACCTTGCAAAAGTGTGAGATGCCCTGGAGAA 1584
      |||||
814 yAlaProAsnAlaGlnGly.....GlnGlyAsnG 824
1585 CTCAAGAAATGACATGCGGTGCCACTCCCGACCG.....CCG 1622
      |||
824 lngln.....ProProProAsnGlyAlaThrPro 833
1623 CGGTCGCCCGCAGCAACCAACAGCAGTACTATGCCAGCCGCTCT...GGA 1669
      |||||
834 ProMetProProAsnGlnlntYrgln..ProAlaProGlylYlAlaProGlnGly 849
1670 TGCCGTCTCGGCCAACAATCCGCTCACCTGTCTTCAACCATTCCT 1714
      |||
850 ProTyrglYlYlYProProProProGlnAlaTyrglYlYProProPro 864

```

Fri Mar 1 09:12:00 2002

us-09-668-119-1.rpt

Page 22


```

|||||
379 GGGCAACCAACCCCTGGAACTCTGGAAATGGCCCTCATGCGATGGCTGT 428
117 ValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAlaL 134
|||||
429 GGTGTACAAACAACCTCAAGCTGAGTGGAGTCCAGCAAGTGGCAT 478
134 eugGlnGlnGlnGlnGlnGlnGlnPheGln.....GlnGlnGlnGln 148
|||||
479 TGCAGCAACAGCAGCAGCGCAACCAACAGCAATTCAGCAACAGCAG 528
149 AlaAlaLeu.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGln 161
|||||
529 GCAGCACTGAGCAACAGCAGCAACAGCAGCAGCAGCAGCAGCAGCA 578
161 nGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnGlnPheGlnAlaVal 178
579 GCAGTTCAGGCAACAGCAATGCTGAGCAACAGTTCAGCAAGCAGTAG 628
178 aGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeu 194
|||||
629 TGCAGCAGCAGCAG.....CTTCAGCAGCAGCAGCAGCAGCAGCTG 672
195 IleLysLeuHisGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 211
673 ATTAAGTTCATCATCAAGCCAGCAACAA...CAGATACACAGCAGCA 719
211 nLeuGlnArgIleAlaGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnG 228
720 ACTGACAGAGATGGACAGTTGCAGCTG.....CAGG 751
228 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 244
752 AACAGCAACAGCAGCAACAGCAGCAGCTTGCAGGCGCCAGCAAGT 801
245 GlnGlnProMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 261
802 CAGCAGCATCATTCAGCAGCAGCAGCAGCCTCCCTTCTCAGGCTAC 851
261 oGlnGlnLeuGlnGlnMetHisIshThrGlnHisIshGlnProPro 278
852 CCAGCAGCTGTCAAGCTGCATCATCCAGCATCAACAGCAGCAGCTC 901
278 InProGlnGlnProProValAlaGlnGlnGlnGlnGlnGlnGlnG 294
902 AGGCTCAGCAGTCCCTCATTTGCTCAAAACCAACACAGCAGATCCA 951
295 GlnSerGlnThrGlnProLeuValSerGlnAlaGlnAlaLeuProG 311
952 CAGTCACAGAGCCAGCTTTGGTGTCAAGCAGCAAGCCCTTCTGAG 1001
311 nMetLeuTyr...ThrGlnProProLeuLysPheValArgAlaPro 327
1002 GATGCTGTATGCTGCCAACAGCAGCTGAATTGTCGTCGTCGATGG 1051
327 aValAlaGlnGlnProProValGlnProGlnValGln..... 338
1052 TGGTCAGCAGCCGCAAGTGCAGGCCCAAGGTCAGAGGTCAGAGCC 1101
339 ...GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 354
1102 GTGACAGCCGAGCAGCAGTGCAGCAGCAGCAGTCTGCCAGATGTA 1151
354 aProGlnVal..... 357
1152 TCCGGGCGTCAGATGATGCTGAAGCTTGGCCCAAGCGGAGTGCAG 1201
357 ..... 357
1202 TAAAGACCCGGTCCCGCCACCTCCACCATGTCTGCTGGCCCTCAAG 1251
358 .....GlnValSerGlnSerLe 364
|||||

```

```

1252 TCATCTCTTTGGGGGAGCAGCCCAACACAGGTCAAGCCAAACAGCT 1301
364 uProMetLeuSerSerProSerProGlnGlnValGlnThrProGlns 381
1302 CACCATGCTGCTTCACCGTCACAGGCGCAGAGTGCAGACCCCAAGT 1351
381 erMetProProProGlnProSerProGlnGlnGlnGlnGlnGlnG 397
1352 CGATCCCACTCCCAACAGCCGTCCTCCCAACACTGGC.....TCA 1392
398 GlnProAsnSerAsnValSerSerGlyProAlaProSerProSerPh 414
1393 CAGCCCAACTGTAATGTCAGCTCCGCGCTGCCCATCTCCACAGCTT 1442
414 eLeuProSerProSerProGlnProSerGlnSerProValThrAla 431
1443 CTGCTGAGCCCTTCACACAGCTTCTCAGAGCCGGGTGACAGCAGCA 1492
431 hrProGlnAsnPhSerValProSerProGlyProLeuAsnThrPro 447
1493 CCCCACAGACTTCAGCGTCTCTCCCTGAGACTTTAAACACCCCTGT 1542
448 AsnProSerSerValMetSerProAlaGlySerSerGlnAlaGln 464
1543 AACCCAGCTGTGTCATGAGCCAGCTGCTCTACAGCTGAGGAGACA 1592
464 nGlnIlyLeuAspLysLeuGlnGlnGlnGlnGlnGlnGlnGlnG 481
1593 GCAGTACCTGGCAAGCTGCAAGCTTGTCCAAATGTCAGACCCCTGC 1642
481 rArgMetIleAsnLysIleAspLysAsnGluAspArgLysAspLeu 497
1643 GACGATGATCAACAAGATCGACAAAGATGAAAGCAAGAAAAAGCA 1692
498 SerLysMetLysSerLeuLeuAspIleLeuThrAspProSerLys 514
1693 AGTAAGATGAAGAGCTGCTGTCATCTCAACCCAGCCCTGGAAGAG 1742
514 sProLeuLysThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 531
1743 TCCCTCGAAGACCTGCAAAAGTGAATGCTGCTGGAAGAGCTCAAG 1792
531 snAspMetArgCysProLeuProHisArgProArgCysHisArgPro 547
1793 ATGACATGGCAGTCCCAACACCCCAACCCCAAGTCTTCTCAACAAA 1842
547 nSerSerThrTyraIaSerArgSerTrpMetProSerTrpProHis 564
1843 CAGCAGAGCAGCTGCCCAACCTCTAGATGCAGTCTGGCCACATCCG 1892
564 IaHisLeuSerSerThrIleProCysThrAlaHisSerPheGlnPro 579
1893 TTCACCTGCTTCAACCAATTCCTGTACCCGACATTTGTGCCAGCA 1939
seq_name: gb_HSU80745
seq_documentation block:
LOCUS HSU80745 1048 bp mRNA PRI 18-DEC-1997
DEFINITION Homo sapiens CTG7a mRNA, partial cds.
ACCESSION U80745
VERSION U80745.1 GI:2565064
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1048)
Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
Breschel T.S., Stine O.C., Callahan C., McInnis M.G. and Ross C.A.
cDNAs with long CAG trinucleotide repeats from human brain
Hum. Genet. 100 (1), 114-122 (1997)
JOURNAL
MEDLINE
97369492
REFERENCE
2 (bases 1 to 1048)

```


REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@gscc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,

Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gun,

Letitia Hsiao, Martin Krzyzanski, Reta Kutsche, Oliver Lee, Soo

Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven

Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline

Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott,

Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,

George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES

Source

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 22 Row: n Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7209306.

Location/Qualifiers

CDS

```
1..2106
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3350171"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NTH_MGC_16"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1..1157
/codon_start=3
/product="unknown (protein for IMAGE:3350171)"
/protein_id="AAH07529.1"
/db_xref="GI:14043091"
/translation="TAVSAIPSSIPYGRPMAYSSSLPMLSSPSPGQOVOTPSM
PPPPSPQSPQSPSSQNSVNSQSPAPSSFLSPSPQSPQSPVTAARTQNSVSPSP
GPLNTVPNPSVMSVMSQSAEQOYIDKLKLSKYIEPLRMINKIDKNEDRKDLIS
KMSKSLDILDFPSKRCPLKTLCKEALKEIKNMAVPTPPPTKQOYLCQPL
DAVLANTRSVPENSLYRTPEVPMATLHGPIITAVVCTRRKRLDEDEROSIPVLOG
EVARLDKRLVNIIDPSPHSCSNNGVTHLCXIDKLDPLSPLEISVSPADYAGSPMLTD
RQWQYDAPFQSYHRCWTKRLDLPKHSVTALLNTMAOSVHQAICLSAA"
BASE COUNT 456 a 719 c 553 g 378 t
ORIGIN
```

alignment_scores: Quality: 1174.00 Length: 226
Ratio: 5.241 Gaps: 1
Percent Similarity: 99.115 Percent Identity: 98.673

alignment block:

US-09-668-119-3 x BC007529 ..

Align seg 1/1 to: BC007529 from: 1 to: 2106

```
355 ProGlyValGlnValSerGlnSerLeuProMetLeuSerProse 371
111 ::::::::::::::::::::::::::::::::::::::::::::::
51 CCCATGCGACAGGTGACGCGAGAGCTCCCATGTCCTCGGCGTC 100
371 ProGlyGlnGlnValGlnThrProGlnSerMetProProProGlnP 388
101 ACCGCGCCACAGGTGACAGCCCGGACGTGATGCCGCCGCCAGC 150
388 ProSerProGlnProGlyGlnProSerSerGlnProAsnSerAsnValSer 404
151 CGTCCCGCGACCGCGCGACCGCGACGCTCAAGCCCAACTCAACGTCAGC 200
405 SerGlyProAlaProSerProSerSerPheLeuProSerProSerProG1 421
```

```
|||||
201 TCTGCGCCCTGGCCCATCTCCAGAGCTTCTGCGCCAGCCCTCACGCGA 250
421 nProSerGlnSerProValThrAlaGthrProGlnAsnPhaserValp 438
251 GCCCTCCAGAGCCAGTACGGCGCGGCGGCCACACAAACTTACGTCTCC 300
438 roSerProGlyProLeuAsnThrProValAsnProSerSerValMetSer 454
301 CCTACCTGACCTTCTTAACACACCTTGAAACCCAGCTTGTCATAGC 350
455 ProAlaGlySerSerGlnAlaGlnGlnGlnGlnThrLeuAsnGlyLeu 471
351 CCACCTGGCTCCACGACGAGCTGAGGACAGACAGTACCTGCAAGCTGAA 400
471 sGlnLeuSerLysTyrIleGluProLeuAlaGlyMetIleAsnLysIleA 488
401 GCACCTGCGAAGTACATCGAGCCCTGCGCCGCGCATGATCAACAAGATCG 450
488 sPlysAsnGluAspArgLysLysAspLeuSerLysMetLysSerLeuLeu 504
451 ACAAGACGAGACAGCAAAAAAGACCTGAGTAAAGATGAAGAGCTTCTG 500
505 AspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 521
501 GACATTCGACAGACACCTCGACGCGGTGTCCTCGAAGACCTGCAAAA 550
521 sCysGlnIleAlaLeuGlnLysLeuLysAsnAspMetArgCysProLeu 537
551 GTGTGAGATTCGCGCTCGAAGAACTCAAGAAATGACATGGCGTCCACTC 600
538 ProHisArgProArgCysHisArgProAsnSerSerThrTyrAlaSerArg 554
601 CCCACCGCCCGCCGCTGCGACGACCAACAGCAGTACCTTGCACGCG 650
554 gSerTrpMetProSerTrpProThrSerAlaHisLeuSerSerThrIleP 571
651 CTCCTGAGATGCGGTCCTGCGCAACATCGCTCACCTGCTTCAACATTC 700
701 CCTGTACCGCAATTCGTCCAGGCCA 726
seq_name: gb_hlg.AC068137
seq_documentation block:
LOCUS AC068137 156315 bp DNA HTG 10-APR-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-803A13, WORKING DRAFT
SEQUENCE 6 unordered pieces.
ACCESSION AC068137 GI:13570009
VERSION AC068137.7
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 156315)
WATERSTON, R. H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston, R. H.
REFERENCE 2 (bases 1 to 156315)
TITLE Direct Submission
JOURNAL Submitted (28-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 10, 2001 this sequence version replaced gi:13431262.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
```



```
403 ..... 403
657 AGGCTGGCGCTGGG:GCCCCCTCCACCCTGGCCCTGGAGGCTGGCCCTG 706
404 .....Ser...SerGlyPro 407
707 CCTGGACCCCGGCTCNCATGTTTCTCCTCCTGGCTGGCTGGCCCT 756
408 AlapProSerSerPheLeuProSerProSerProGlnProSerG1 424
757 GCCCATCTCCAGTACCTCTGCTGCCAGCCCTCACCAGCCAGCCCTCCA 806
424 nserProValThrAlaArgThrProGlnAsnPheSerValProSerProg 441
807 GAGCCAGTGAAGGCGCGAGCCCGACAGACTTCACTGCTCCCTCACCCTG 856
441 LProLeuAsnThrProVal..... 447
857 GACCTTTAAACACACACTGCTGAAGTTGGGCTGAGGCTGCTAAGTCACTCC 906
447 ..... 447
907 TCACCTTTATGAGGCTTCAGCTCATCTGGGTGCGAGCTGGGGGCC 956
447 ..... 447
957 TCAGAGCTCAAGTTCCACCCCGAGGGTGGAGGCTGTGGCTCACCCTG 1006
448 .....LysProSerSerValMetSerProAlaGlySer 459
1007 CTGTCCTCGAGTG:ACCCAGCTCTCTCATGAGCCAGCTGCTCCAG 1056
459 rGlnAlaGlnGlnGlnGlnTyLeuAspLeuLeuLeuLeuLeuLeuLeuLeu 476
1057 CCAGGCTGAGGAGCAGCAGTACCTGAGCACTGAGAGCTGTGCAAGT 1106
476 rTleGlnProLeuAsnGlnMetLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 492
1107 ACATGAGACCCCTGCTCCGATGATCAACAGATCGACAGAGAGAGAGAGC 1156
493 ArgTyrLeuAspLeuSerLeuMetLeuSerLeuLeuLeuLeuLeuLeuLeu 509
1157 AGAATAAGAGAGCTGTGTAGATGAAGAGCCTCTGACATTCGACAGA 1206
509 pProSer..... 511
1207 CCCCCTGAAGGGTGTGCTTTGCCACAGCCACGAGGAGTCACAAAGG 1256
511 ..... 511
1257 CACAGATAGCCCAAGCATGATGGGCACTTGATGATGATGATGATGATGAT 1306
511 ..... 511
1307 AAAGGACACGAGCAGCTCTTTGACCCCTGGCCAGAGGCTTCACAGGCTCC 1356
511 ..... 511
1357 ACTGTGTGTGTGTGTGGGCTTCAAGCCAGGCTTCATCTTGGCCATGC 1406
511 ..... 511
1407 CCAGCTTTGGCCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1456
511 ..... 511
1457 TCACAGCCAGACCTCATCCAGTCAGCAGCAGGAGGAGGAGGAGGAGGAGT 1506
511 ..... 511
1507 CCTGACCATCAGCTGGCCAGATGGGCTGAGCCTGACCTGAGAGTTCTGC 1556
511 ..... 511

511 ..... 511
1557 CCTGACTTGTGTGATGACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1606
511 ..... 511
1607 TTTGTAGCCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1656
511 ..... 511
1657 AGAAGCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1706
511 ..... 511
1707 GAGCTGGGAGGAGTACCTCCAACTGCATGGGAGTGTGATATACCTCCCT 1756
511 ..... 511
1757 TTACCTCAACTAGGCTATCCCCCAACCCCTGTGTAGAGGAGTCTCT 1806
511 ..... 511
1807 GTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1856
511 ..... 511
1857 TCTTCATGGGTGATCCTTTGTTAGGGTTATAGGATGAGTGGGAGGAGG 1906
511 ..... 511
1907 ACTGGCTACATGACATGAGGAGGATTCACAGGCTGATCAGGGGCTGGGC 1956
511 ..... 511
1957 ACTGTCCCTTCTCTTACCCATCACCCTCCTGCTGCAACGAGGAGAGC 2006
511 ..... 511
2007 AGCTCTGGGGGTGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2056
511 ..... 511
2057 CTGTCACTGATGAGGAGGCTCTGAGAGGTTTGTGATGCTGAGGAGCCA 2106
511 ..... 511
2107 GTGGCAGTAGTTGGATCCTGAGCCTGAGAGCTGACATGTCTTCCACTCT 2156
512 .....Lys...ArgCysProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 521
2157 TTCAGCCCAAGTCTCTCTCTCTGAGGAGTGTCCCTGAAGACTTGCATA 2206
521 yscysGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 534
2207 AGGTGTAGATGCGCCCTGAGAGAACTCAAGATGACATGGCGGTGTGAGT 2256
534 ..... 534
2257 GGGATCCAGACACCCCTAGGGGAACAGAGGCTCTCTTAAGACTCTG 2306
534 ..... 534
2307 GAGTGGCGGTGAGAGGCTTCAAGGTGAGGAGTGGGCGGAGGAGGAGG 2356
534 ..... 534
2357 GCGTACCTTGGACCCCTGCCAGAGGCTTCTGCGCAGGTCTGTGTGTC 2406
534 ..... 534
2407 TGGGTGGAACATGGAGAGAGTCAACCTCTGTCTACGAGGAGGAGGAGGAG 2456
534 ..... 534
```



```

2457 CCTCCTTACCCATTTCAGAGAGGAGGAGGACTGTAGCGAGATAGGCTG 2506
534 ..... 534
2507 TGGGGGTGGGGGATTAATTCAGGATCAGCCACATTTGTCTGCACAAGG 2556
534 ..... 534
2557 TGGAGGCTGTAGAGAGGCGCTCGGAATCTGACTGTAGTGACCAATGGGC 2606
534 ..... 534
2607 CTGGGGGTGAAGGCCCCCTAAATGGGGAACCTCGGGTCCGGCTGCT 2656
535 ..... CysProLeuProHisArgProArgCysH1 544
2657 GACCGTCCCATCTGTCCTCAGCCACCTCCCGCCGCGTGCCA 2706
544 SATGPROASerSerThrTyraIAserArgSerTrpMetProSerTrp 561
2707 CCGACCAAGAGCACTATGCTCCACCCGCTCTGGATCCCTCTGGC 2756
561 roHrSerAlaHisIeuserSerThrIleProCysThrAlaHisSerPhe 577
2757 CAACATCGCTCAGCTCTCTTCAACCACTTCCTGTACCGCACATCTGTC 2806
578 GlnPro 579
2807 CAGCCA 2812

seq_name: gb.Pr.AC004033
seq_documentation_block:
LOCUS AC004033 145356 bp DNA PRI 31-MAY-2000
DEFINITION Homo sapiens Chromosome 22q11.2 PAC Clone p_m11 In BCR12-GGT
ACCESSION AC004033
VERSION AC004033.3 GI:4581183
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Budarf,M.L. and Emanuel,B.S.
JOURNAL Unpublished
2 (bases 1 to 145356)
Zhang,G., Lao,V., Zhan,M., Budarf,M.L. and Roe,B.A.
REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L. and Roe,B.A.
JOURNAL Direct Submission
Submitted (26-JAN-1998) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 145356)
Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
JOURNAL Direct Submission
Submitted (09-DEC-1998) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 145356)
Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
JOURNAL Direct Submission
Submitted (26-JAN-1999) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 145356)
Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
JOURNAL Direct Submission
Submitted (04-FEB-1999) Department of Chemistry And Biochemistry,

```

```

REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
JOURNAL Direct Submission
Submitted (09-APR-1999) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
8 (bases 1 to 145356)
Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
JOURNAL Direct Submission
Submitted (12-APR-1999) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
9 (bases 1 to 145356)
Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
REFERENCE
AUTHORS Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
JOURNAL Direct Submission
Submitted (31-MAY-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Apr 12, 1999 this sequence version replaced gi:4580479.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC005500(p52f6) 112352 192592 (0) overlaps AC004033(p_m11) 1 78509
(66847) AC004033(p_m11) 89227 145356 (0) overlaps AC007050(bac32) 1
59304 (104604).

FEATURES
Source
1..145356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/chromosome="22q11.2"
/map="22q11.2"
/clone="p_m11"

BASE COUNT 34453 a 36720 c 37358 g 36825 t
ORIGIN

alignment_scores:
Quality: 765.50 Length: 949
Ratio: 2.613 Gaps: 19
Percent Similarity: 30.875 Percent Identity: 26.238

alignment_block:
US-09-668-119-3 x AC004033 ..
Align seg 1/1 to: AC004033 from: 1 to: 145356
228 GlnGlnGlnGlnGlnGlnGlnGlnAlaLeuGlnAlaGlnProPro11 244
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
119558 CAGAGCAGCATGTGCACAGCAGCAGCGGTGCAGAGGAGGAGGCGTCA 119607
244 eGlnGlnProPrometGlnGlnProGlnProProSerGlnAlaLeuP 261
|||||
119608 GTGCCAGCCA.....AGGCGATTGC 119627
261 roGlnGlnLeuGlnGlnMetHisThrGlnHisGlnProProPro 277
|||:|||||
119628 CTCCAGACTGCCCTTTCTCCTCAGACTC.....CAGCAGCAGCCCTC 119671
278 GlnProGlnGlnPro.....ProValAlaGlnGlnProse 290
|||:|||||
119672 AAGGCTGTGCCCTGTGGTCCCTGCGCAAGAGGAGGCTGCCCTGC 119721
290 rGlnLeuProGlnGlnSerGlnThrGlnProLeuValSerGlnAlaGlnA 307
|||:|||||
119722 CTGGGCCCGCCCGCCAGC.....CCCTGTGTCACTCATGC...T 119759
307 laLeuProGlnGlnMetLeuTyThrGlnProProLeuLysPheValArg 323
|||:|||||
119760 GTGGCCAGGC.....TGGGGAGGT 119779

```

```
334 AlaProMetValVal3InglInProValGInPro..... 335
:::
119780 GTGCTGTGTATGTCAGGTGATTTGGATGAAAGACAGAGGTGTGTG 119829
336 .....GlnValGInG 339
119830 GTGAGCTGTGGGGAGGATGGGCTATGCGGTGGGAGGTGTGT 119879
339 InGInlInThAlaVal1GInlInThAlaInlAlaInlMetValAlaPro 355
:::
119880 GGAGCGAGCGGGCGGCGCTGTG.....GCAGCGCGCGGCTGAGCCCC 119923
356 GlyVal..GlnVal3ArgInSerLeuPrometLeuSerSerProSer 371
:::
119924 TCCACTTCCAGGTCAACCCAGACAGACCTCCCATGCTGTCTCCGCGTCA 119973
372 ProGlyInGInlVal3InlInProInSerMetProProProGInlPr 388
:::
119974 CCGGCGCAGCGAGTGAAGACCCGAGTCAGTGCCTCCGCCAGCC 120023
388 OSeProGInlProGlyGInProSerSerGInProAsnSerAsnVal... 403
:::
120024 GTCCCGCAGCCCGGACAGCCAGCTCAACAGCTCAGGT 120073
403 ..... 403
120074 AGGCTGTGGCTGGGGTGCCTCCCACTGGCCCTGAGGCTGGCCCTG 120123
404 .....SerGlyPro 407
120124 CCGGACCCCGGCTC/CATGTTCTCTCATCTTGGCTGAGCTGTGGCCCT 120173
408 AlaProSerProSerSerPheLeuProSerProSerProGInlProSerG 424
:::
120174 GCCCATCTCCAGT/GCTTCTCCAGCCCTCACCGGAGCCCTCCCA 120223
424 nSerProValThAlaArgThProGInlAsnPhSerValProSerProG 441
:::
120224 GAGCCGAGTGGCGCGGACCCAGAACTTCAGTCTCCCTCACCTG 120273
441 LysProLeuAsnThPro..Val..... 447
:::
120274 GACCTTTAAACACACCTGTGTAAGTGGGCTGAGGTGCTAAGTCACTCC 120323
447 ..... 447
120324 TCACCTTTATGAGGCTTCAGCTCACTAGGTGTGCGAGCTGTGGGCC 120373
447 ..... 447
120374 TCAGAGCTCAAGTTCCCAACCCAGGCTGAGGCGTGTGCTCACCCGC 120423
448 .....AsnProSerSerValMetSerProAlaGlySerSe 459
120424 CTGTGCTCTGCAGTGAACCCAGCTGTGTCATAGACCCAGCTGTGCTCAG 120473
459 rGlnAlaGlnGlnGlnGlnTyrlLeuAspLysLysGlnLeuSerLysT 476
:::
120474 CCAAGGCTGAGGAGACCCAGTACCTGGACAGCTGAAGCAGCTGTCAAGT 120523
476 yrlleGlnProLeuArgArgMetLeuAsnLysLysLysAsn..Glu.. 491
:::
120524 ACATCGAGCCCTGCGCCGATGATCAAAAGATCGACAAGAAAGAAAGT 120573
491 ..... 491
120574 AGGCTGACGACGAGGCGCTGCACCTGGGACACACACAGCTTG 120623
492 .....AspArgLysLysAspLeuSer 498
120624 TGCTTAGTGTATCTCTTCTGTGTCCCGACAGAAAAAGACGTGAGT 120673
499 LysMetLysSerLeuLeuAspLysLeuThrAspProSer..... 511
:::
120674 AAGATAGAGGCTTCTTGAGCATTTGTGACAGACCCCTCGAAGCGGTGAGC 120723
511 ..... 511
120724 TTTGCCACAGCCCGACGAGGCTCCACAAGGCGACAGATAGCCAGCCAT 120773
511 ..... 511
120774 GGATGGGCACTGTGTATGATGTGGTTTAAACAAGGACAGGACGCTC 120823
511 ..... 511
120824 TTTGACCTGGCGCAGAGAGGCTCCAAAGCTCAGTGTGTGTGTGG 120873
511 ..... 511
120874 GCTTCAAGCCAGGCTTCACTTGGCCCATGCCAGCCTTGGCCATTC 120923
511 ..... 511
120924 CTGGCTGTGTGTGTGCTCATGTGTGGCCATCAAGACGACGACCTCATC 120973
511 ..... 511
120974 CAGTCAGCAGCAGGCGCCAGCTTGCACTGGGCTGACATGAGCTGGGC 121023
511 ..... 511
121024 CAGATGGGCTGAGCTGACCTGAGTTCGAGTTCGCCCTGACTTCTGTGTAC 121073
511 ..... 511
121074 CATGGGAGGCACTGATCTAGTTACCCATTGTGACCCAGCTGTGT 121123
511 ..... 511
121124 TGTCAATGAATAATGAACAAGGTGGCTGTGTGAGAAGAGGTGAAGCA 121173
511 ..... 511
121174 GGGCTGTGGCCAGAGGCTGTGTGAGGATAGAGCTGGGCAATACCTC 121223
511 ..... 511
121224 CAAACTGATGGGCAATGGCATATATACCTTCCACTCAACTCAGGCT 121273
511 ..... 511
121274 ATCCCCCAACCCCTGTAGAAGGGTCTCTGTCTGTCTGTGTGATA 121323
511 ..... 511
121324 GAATCGCATCTATATGGGTGTCTCCAGGTCTTCAATGGGTTGATCT 121373
511 ..... 511
121374 TGTGTAGGGTTATATAGGATGAGTGGGAGGAGGCTGCTGATGACAT 121423
511 ..... 511
121424 GAGGCAATTCACAGCTGATCAAGGGGCTGGCACTGTCTCTTCTTGAC 121473
511 ..... 511
121474 CCATCACTCACTTGCCACACGAGGACAGACAGCTCTGGGGTTGAGG 121523
511 ..... 511
121524 CTGCTGCTTCAACTGTGGTGTGGGAACAGCTGTCACTGTGTAGAG 121573
511 ..... 511
```

[illegible]

JOURNAL REFERENCE AUTHORS TITLE JOURNAL	Unpublished 2 (bases 1 to 163908) Chen,F., Hu,X., Emanuel,B., Budarf,M. and Roe,B.A. Direct Submission Submitted (09-MAR-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	3 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (07-JUL-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	4 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (19-JUL-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	5 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (22-JUL-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	6 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (29-JUL-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	7 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (26-AUG-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	8 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (01-SEP-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	9 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (19-OCT-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	10 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (20-NOV-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	11 (bases 1 to 163908) Chen,F., Hu,X., Budarf,M., Emanuel,B. and Roe,B.A. Direct Submission Submitted (31-MAY-2000) Department of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT	On Nov 20, 1999 this sequence version replaced gi:6067157. Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below. AC004033(p.m11) 89227 145356 (0) overlaps AC007050(dac32) 1 59304 (104604) AC007050(dac32) 141234 163908 (0) overlaps AC007308(pac408) 1 22675 (193113). Location/Qualifiers 1. 163908
FEATURES SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22q11"


```

511 ..... 511
31894 TGTATGAAATGAAACAGGTGGCGCTGGTGAGAACAGGTGGAAAGCA 31943
511 ..... 511
31944 GGGCTGCTGGCCACAGGCTGCTGTGAGAGTACAGCTGGGGCAGTACCTC 31993
511 ..... 511
32044 ATCCCCCAACCCCTGTCTAGAAAGGCTCTCTGTCTGTCTGTCTGTGATA 32093
511 ..... 511
32094 GAATGCCAGTCACTATGGGTGCTTCAGAGTCTTATGGGTGATCCT 32143
511 ..... 511
32144 TTGTTAGGTTATAGGATGAGTGGGGCAGGAGCTGGCTTACATGACAT 32193
511 ..... 511
32194 GAGGATTCACAGAGCTGTGATCAGGGGCTGGCCACTGTCCTTCTTTGAC 32243
511 ..... 511
32244 CCATCACTCACTCTGCCAACCCAGGACAGCAGCTCCTGGGTTGAGGG 32293
511 ..... 511
32294 CTGCCTGCTTCACTCTGGGTCTGGGAACAGCTGTCAACGTGTAGAG 32343
511 ..... 511
32344 CTCCTGACAGGTTTCTGATGGCTGAGGCCAGTGGCAGTACTGGGATC 32393
511 ..... 511
32394 CTGAGCCTGAGACTGCTCTTCCACTCTTTCAGCCCAAGCTCCTC 32443
512 ..Lys..ArgCysProLeuLysThrLeuGlnLysCysGlnIleAlaLeuG 527
::: |||||
32444 TCTCTGACAGTGTCCCTGAAAGACCTTGCAAAAGTGTGAGATGCCCTGG 32493
527 LulysLeuLysAsnAspMet Arg..... 534
32494 AGAACTCAAGATGACATGGCGGTGTGATGGATGCCACACACCCCT 32543
534 ..... 534
32544 AGGGAAACAGGGCTCTCTAAGAGCTCTGGAGTGTGTGAGAGGCG 32593
534 ..... 534
32594 CTTCAAGTTCAGGATCTGGGGGGGGGGCGGCTGACCTTGGACCTTG 32643
534 ..... 534
32644 CCCACAGAGCTTCTGCTGGCCAGCTGTGCTGGTGGGGAACATGGAG 32693
534 ..... 534
32694 AAGTCACTCTGTCTAGAGCCCCCGTGGGTGCTCTTACCCATTGCA 32743
534 ..... 534
32744 GAGAGGCGAGGACTGTAGGAGATAGGCTGTGGGGGTGGGATTTAT 32793

```

```

534 ..... 534
32794 CCAGGATACGCCACATTGTCTGCACAAGGCTGGAGGCTGTGAGAGG 32843
534 ..... 534
32844 CCTCGGAATCTGACTGTGATGATGACCATGGCCGTGGGTGTGAAAGCCCC 32893
534 ..... 534
32894 CTAAATGGGAAACCTTCGGGTCCCGGGCTGTCAGCCGTCCTCATCTC 32943
535 ...CysProLeuProHisArgProAlaArgCysHisArgProAsnSerThr 550
|||||
32944 TCCAGGCCACTCCCGCCACCGCCCGGTGTCACACGCAAAAGCAGTAC 32993
550 rTyraIAserArgSerTrpMetProSerTrpProThrSerAlaHisLeuS 567
|||||
32994 CTATGCCAGCCGCTCTGGATGCCGTCTGGGCCACATCCGCTCACCTGT 33043
567 eSerThrIleProCysThrAlaHisSerPheGlnPro 579
|||||
33044 CTTCAACCAATTCCTGTACCGCACATTCGTTCCAGCCA 33081
seq_name: gb_hhg:AL358856
seq_documentation block:
LOCUS AL358856 157080 bp DNA HTG 10-JUN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-4G23, *** SEQUENCING IN
PROGRESS ***, 12 unordered pieces.
ACCESSION AL358856
VERSION AL358856.3 GI:9863846
KEYWORDS HTG: HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Mclay,K.
Submitted (09-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:9214244.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA4623
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: 152367 bases at least Q40
Consensus quality: 15799 bases at least Q30
Consensus quality: 154743 bases at least Q20
Insert size: 155980; sum-of-contigs
Insert size: 165281; 6.3% error; agarose-fp
Quality coverage: 5.20x in Q20 bases; sum-of-contigs quality
coverage: 4.91x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3463: contig of 3463 bp in length
* 3464 3563: gap of 100 bp

```

```
* 3564 13200: contig of 9637 bp in length
* 13201 13300: gap of 100 bp in length
* 13301 59189: contig of 45889 bp in length
* 59190 59289: gap of 100 bp in length
* 59290 85824: contig of 26535 bp in length
* 85825 85925: gap of 100 bp in length
* 85926 88142: contig of 2218 bp in length
* 88143 88242: gap of 100 bp in length
* 88243 95034: contig of 6791 bp in length
* 95035 100091: contig of 4958 bp in length
* 100092 100191: gap of 100 bp in length
* 100192 117326: contig of 17135 bp in length
* 117327 117426: gap of 100 bp in length
* 117427 119946: contig of 2520 bp in length
* 119947 120046: gap of 100 bp in length
* 120047 130509: contig of 10463 bp in length
* 130510 130609: gap of 100 bp in length
* 130610 139309: contig of 8700 bp in length
* 139310 139409: gap of 100 bp in length
* 139410 157080: contig of 17671 bp in length.
Location/Qualifiers
1. 157080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-4G23"
/clone_1lb="RP11-1.1"
1. 3563
/note="assembly-fragment:00230
clone_end:SP6
vector_side:left"
misc_feature
/note="assembly-fragment:00057
fragment_chain:1"
13301. 59189
/note="assembly-fragment:01204
fragment_chain:1"
59230. 85824
/note="assembly-fragment:01664
fragment_chain:1"
85925. 88142
/note="assembly-fragment:00505
fragment_chain:1"
88243. 95034
/note="assembly-fragment:01296
fragment_chain:2"
95134. 100091
/note="assembly-fragment:01408
fragment_chain:2"
100192. 117326
/note="assembly-fragment:00869
fragment_chain:2"
117327. 119946
/note="assembly-fragment:00628
fragment_chain:3"
120047. 130509
/note="assembly-fragment:01620
fragment_chain:3"
130610. 139309
/note="assembly-fragment:01199"
139410. 157080
/note="assembly-fragment:01779"
BASE COUNT 37302 a 35841 c 39467 g 39369 t 1101 others
ORIGIN
```

```
FEATURES
source
misc_feature
/note="assembly-fragment:00230
clone_end:SP6
vector_side:left"
misc_feature
/note="assembly-fragment:00057
fragment_chain:1"
13301. 59189
/note="assembly-fragment:01204
fragment_chain:1"
59230. 85824
/note="assembly-fragment:01664
fragment_chain:1"
85925. 88142
/note="assembly-fragment:00505
fragment_chain:1"
88243. 95034
/note="assembly-fragment:01296
fragment_chain:2"
95134. 100091
/note="assembly-fragment:01408
fragment_chain:2"
100192. 117326
/note="assembly-fragment:00869
fragment_chain:2"
117327. 119946
/note="assembly-fragment:00628
fragment_chain:3"
120047. 130509
/note="assembly-fragment:01620
fragment_chain:3"
130610. 139309
/note="assembly-fragment:01199"
139410. 157080
/note="assembly-fragment:01779"
BASE COUNT 37302 a 35841 c 39467 g 39369 t 1101 others
ORIGIN
```

```
alignment_scores:
Quality: 705.50
Ratio: 2.486
Percent Similarity: 27.805
Length: 1025
Gaps: 11
Percent Identity: 23.415
alignment_block:
```

US-09-668-119-3 x AL358856/rev ..

Align seq 1/1 to reverse of: AL358856 from: 1 to: 157080

```
39 HispEarGasp1LeHisAnLysSergL..... 49
:::|||||
77750 TATTATAGACTAATCATGATAGGAAATCCCAAAATCAACAAATGCTA 77701
50 .....AlaSerValSerAspPromLeuAlaLeuGlnSerL 62
|||||
77700 CTGCTATCTATCTTCTCTCTGTGTAAATGCAAGCTTCAGATT.... 77655
62 eutHrGlyGlyProAlaAlaGlyAlaAlaGlyLeuGlyMetPropArg 78
|||||
77654 .....CCAGCGCA..GCCGTGGCACTGGGGCTTTG..... 77625
79 GlyProGlyGlnSerLeuGlyGlyMetGlySerPheGlyAlaMetGlyL 95
:::|||||
77624 .....CTGGCTGTTTGTCTCCCTGCTGTGCA 77599
95 nPromSerLeuSerGlyGlnProProGlyThrSerGlyMetAlaP 112
|||||
77598 GCCCTGCAGCGTTTCTGGGAATCTGCCCTGTGCACTGCTGCA..... 77556
112 rHisSerMetAlaValAlaSerThrAlaThrProGlnThrGlnLeuGln 128
:::|||||
77555 .....GACTGTGCTTTTCTCACGCCAGCTGCAG 77526
129 LeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGlnPheGlnL 145
|||||
77525 CTCACAGAGTGGCGCTGCAGCAGCAGCAGCAACACAGCATTCACAGA 77476
145 ngInGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 162
|||||
77475 GCAGCAGCAGCGCGCTACAGACAGCAGCAGCAGCAGCAGCAGCAGC 77426
162 InPheGlnAlaGlnGlnSerAlaMetGlnGlnGlnPheGlnAlaVal 178
|||||
77425 AGTTCAGGCTCAGCAGAGTGCATGACAGCAGCAGCAGCAGCAGTGTG 77376
179 GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnL 195
|||||
77375 CAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCATTA 77326
195 eLysLeuHisHisGlnAsnGlnGln..... 203
|||||
77325 TAAATTGCATCATCAAAATCACCAACAGTACAGTCCCTGTTCTGTC 77276
203 ..... 203
77275 TCTTGGCTCCCTCTGCAGCGTGAAGCCTGGCTGGCATGACCAAT 77226
203 ..... 203
77225 GCTGGTGCAGTGCAGGAGCAGCAGCGAGGTCTTGTGACAGGTGGCC 77176
203 ..... 203
77175 TGCCCTTTCCATGGGCTTCTCAAAAAGCTTGAGCAAGCGCGGCCAG 77126
203 ..... 203
77125 TGGCTATGCTGTATATCCAGCATTGGGAGGCCAAGCAGCGGATC 77076
203 ..... 203
77075 ATCAGAGTTCAAGACAGCCTGGCCAGCATGTGGAACCCCATCTCTAC 77026
203 ..... 203
77025 TAAATAATACAAAATTAAGCTGGAGCATGTGGCTACACTATATCCCA 76976
203 ..... 203
```

76975 GCTATTGGAGGCTGAGCGAGAGAACTGATGAACCTGGGAGCGGAG 76926
203 203
76925 GTTGCAGTAGCCGAGATGCGCGCTGCGCTCCAGCCTGGCAGCAGCG 76876
203 203
76875 AGACTCTGTCTCAAAAAATAAATAAAGAAAGCAAAAGTCTGG 76826
203 203
76825 GACAGCGAGAGAAACAGCCCTGACTACTAGAAAGTTCTGTATGCA 76776
203 203
76775 GCCCTTCATTGGGCTTGGGCATCTGGCGCTGGCCCTCCTGTAGTTG 76726
203 203
76725 AGGGCTGCGAGGAAATCAAAACAAGATGAGTGAAGCAAGATGAGAC 76676
203 203
76675 GAACACTCTGAGGCTTGGACTACTGAGCCATTGGGCCATTGGCT 76626
203 203
76625 GAGSACACATGTGCTCCTTTACCTGTGTCTGGTTCTGGGCTCA 76576
203 203
76575 GTGTAGTCGACCTCCAGCTGGTGTTCATGAGGTGATGTCATG 76526
203 203
76525 TTAGCAGACTGCTTCAGAAAGTTTCACTTCCATGTTCCATAG 76476
203 203
76475 AGAGTCTAGAGTGTAAACATCTGCCCTCAGCAGAGCTGTGTTGGA 76426
203 203
76425 GAGSAAAGCCCAAGGCCCCACCCTCCACCCCTTCAGGCTCAG 76376
203 203
76375 CTAGCTGGGCTTCTGTGATCTTGACACCCCTTCCAGAGCTGACATG 76326
203 203
76325 CATCTGCAGCCACATTTCTCTCTCTGCTGACAGACAGACCTTGACCG 76276
203 203
76275 TCACCGTAGGCTCATGCCACTGACCTATAGGGCAAGACATGCTT 76226
203 203
76225 GCAGAGCTTCAGAAATCTGCAGTTGCCCAAGTTTATCTTACCTTCA 76176
203 203
76175 TGGAGCTTACCTATCTCTGTATCTTTGCTGAAGTTCTAGATGTC 76126
203 203
76125 CCTTGAACCATCTAGTGAATGTAGACGAGCCAGCAGACACTT 76076
203 203

76075 GGTGCCAGGCAACACTTGTGCTCCCTCCAGTGTACGAAAGCAGCTGT 76026
203 203
76025 GTTGGCTCGGGGGGCCAGAGCCCTTCTGTGGAGATGCCCCCCTCCC 75976
203 203
75975 GTGTAGTAGAGGTACAGCTGTACAGCCAGAGAGCCCTTTTTTTTTT 75926
203 203
75925 TTATTGAGACTTTTTTTGCCCTTACTCTGTGTGCCAGCTGAGATGCA 75876
203 203
75875 GTGTGTATCTCGGCTCAGTGCACACTGTGACCCACAGTTTAAGTAT 75826
203 203
75825 TCTCCTGCCTCAGCCTCCCAAGTACGTGGGATTACAGGTGAGCCACCA 75776
203 203
75775 TGCTGGCTAATTTTGTATATATAGTAGAGACGGGTTTGGCCAGTTG 75726
203 203
75725 TCCAGCTGTCTCGAATCTGTACCTCAGTGATCTGCCCGCTCGGCC 75676
203 203
75675 TCCCAAGTCTGGGATTACAGGCGTAGGCCACTGCACCCAGCAGAA 75626
203 203
75625 GCTTCTGTATGAGCTTCGACATGGCCACTGGGCTTCTGAGGCCAGA 75576
203 203
75575 CTGTACAGGAGAGAGCTGTGCCACAGAGCAGATGATGATTAACCGAG 75526
203 203
75525 TGCTGCTGTCTGTCTAGATACAGCAGCAGCAACACACTCAGAGG 75476
203 203
75475 AATAGCAGAGCTGACCTCCCAACACAGCAGCAGCAGCAGCAGCAGC 75426
203 203
75425 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 75376
203 203
75375 CCGATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 75326
203 203
75325 uGlnGlnMetHisHisThrGlnHisGlnProProGlnProGln 281
264 203
75281 InProProValAlaGlnAsnGlnProSerGlnLeuProGlnSerGln 297
264 203
75297 AGCTCCAGTTGCTCAGAACCAACCACTCCCGCAGCTGAGCAG 75226
203 203
75225 ACCAGGCTTTGGTGCACAGGCGCAAGCTCTCCGGAAGTGTGTA 75176
203 203
75175 TACCAACCACTGAAATTTGTG...AGTACTGTGGCCACAGTGA 75129

```

331 roProValGlnProGlnValGlnGlnGlnInThrAlaValGlnThrAla 347
      ::||| ::||| ::||| ::|||
75128 GCACATGACGACCCGTACCTCTGTGACAGATATTCTAGGCTTTGGCT 75079
      ::||| ::||| ::||| ::|||
346 GlnAlaIaIaGlnMetValAlaProGlnValGlnValSerGlnSerLe 364
      ::||| ::||| ::||| ::|||
75078 AAGATAGATATCTTCTTCAAGTGTGACACTCAGGACGCCCA 75029
      ::||| ::||| ::||| ::|||
364 uProMetLeuSerProSer..... 371
      ::||| ::||| ::||| ::|||
75028 CCCCTTGCCAGCCCTTCCGACTCTAGATGGCTCAGAGGAGCATGTG 74979
      ::||| ::||| ::||| ::|||
372 .....ProGlnGlnVal 376
      ::||| ::||| ::||| ::|||
74978 TTTTCATTGTGTGTGTGAGAGCTTGTGACCCCGCACAGGCTACT 74929
      ::||| ::||| ::||| ::|||
377 GlnThrProGlnSerMetProProProProGlnProSerProGlnProG 393
      ::||| ::||| ::||| ::|||
74928 GCTGCTGGTGGGACCA.....AGTCG 74903
      ::||| ::||| ::||| ::|||
393 yGlnProSerGlnProAsnSerAsnValSerSeryl..... 406
      ::||| ::||| ::||| ::|||
74902 CATTCAGACAGCATGTGCTGATGACGACATTTGTGGGCCCTTCACGG 74853
      ::||| ::||| ::||| ::|||
407 .....ProAlaPro 409
      ::||| ::||| ::||| ::|||
74852 TGTGACACTTCGACATGATGACACGACATGAAGATGTGACGCTCCG 74803
      ::||| ::||| ::||| ::|||
410 SerProSerSerPheLeuProSer 417
      ::||| ::||| ::||| ::|||
74802 TCAGTGTCTGCTGTGCTTGCACACT 74779
      ::||| ::||| ::||| ::|||
seq_name: gb_hlg:AC087802;

seq documentation block:
LOCUS AC087802 187315 bp DNA HTG 23-APR-2001
DEFINITION Mus musculus chromosome 16 clone rp23-472f15, WORKING DRAFT
SEQUENCE 4 ordered pieces.
ACCESSION AC087802.5 (1:13752732
VERSION AC087802.5 (1:13752732
KEYWORDS HTG; HTGS_PHS2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187315)
Mus musculus Chromosome 16 BAC Clone rp23-472f15
2 (bases 1 to 187315)
Jiang, X., Sorg, L. and Roe, B.A.
Unpublished
Submitted (26-JAN-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Apr 23, 2001 this sequence version replaced gi:12740049.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 7409 7508: contig of 7408 bp in length
* 7509 17332: contig of 9624 bp in length
* 17333 17432: gap of unknown length
* 17433 93194: contig of 75762 bp in length
* 93195 93294: gap of unknown length
* 93295 187315: contig of 94021 bp in length.

```

```

FEATURES
    source          Location/Qualifiers
                    1..187315
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /chromosome="16"
                    /clone="rp23-472f15"
                    /colone_11b="rp23"
BASE COUNT      49420 a 44136 c 42281 g 51159 t      319 others
ORIGIN

alignment_scores:
    Quality: 679.00      Length: 457
    Ratio: 3.031      Gaps: 15
    Percent Similarity: 49.015      Percent Identity: 41.357

alignment_block:
US-09-668-119-3 x AC087802/rev ..

Align seg 1/1 to reverse of: AC087802 from: 1 to: 187315

276 ProProGlnProGlnInProProValAlaGlnAsnGlnProSerGlnLe 292
      ::||| ::||| ::||| ::|||
55179 CCGAGCTCCCTCCTTATACACAGAGTGTGAGAGACAGGCATTAACAT 55130
      ::||| ::||| ::||| ::|||
292 upro...ProGlnSerGlnThrGlnProLeuValSerGlnAlaIaIaL 308
      ::||| ::||| ::||| ::|||
55129 ATCTATACACAGCCCAAGAAACCTGTTGCTGCTCCACCAA.... 55084
      ::||| ::||| ::||| ::|||
308 euProGlnGlnMetLeuTyThrGlnPro.....ProLeuLysPhe 321
      ::||| ::||| ::||| ::|||
55083 .....GGAGCAGGCTACCTACCTGCTGCTTCTTCTTACGCTTCGATAC 55039
      ::||| ::||| ::||| ::|||
322 ValArgAlaProMetValAlaGlnInProProValGlnProGlnValG 338
      ::||| ::||| ::||| ::|||
55038 .....CATGTGTATGTGCTGCCAGAGGTAGCGTGTGCTGCTGAGTAA 54995
      ::||| ::||| ::||| ::|||
338 ngInGlnGlnInThrAlaValGlnInThrAlaGlnAlaGlnMetValAla 355
      ::||| ::||| ::||| ::|||
54994 AGAAATPAGGGTGGATPAGATGATGATGATGATGATGATGATGATGAT 54945
      ::||| ::||| ::||| ::|||
355 roGlyVal.....GlnValSer 360
      ::||| ::||| ::||| ::|||
54944 AGGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 54895
      ::||| ::||| ::||| ::|||
361 GlnSerSerLeuProMetLeuSerSerProSerProGlnGlnValG 377
      ::||| ::||| ::||| ::|||
54894 CAAAGAGCCTCACCATGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 54845
      ::||| ::||| ::||| ::|||
377 nThrProGlnSerMetProProProProGlnProSerProGlnProG 394
      ::||| ::||| ::||| ::|||
54844 GACCCACAGTGTGATGCCACCTCCCAAGCCGCTCCCAACACCTGAC. 54796
      ::||| ::||| ::||| ::|||
394 InProSerSerGlnProAsnSerAsn..... 402
      ::||| ::||| ::||| ::|||
54795 .....TCACAGCCCAACTATATGTCAGTAGGCTGTGGGATGAC 54754
      ::||| ::||| ::||| ::|||
402 ..... 402
      ::||| ::||| ::||| ::|||
54753 CCTTCCCTAGCCTTCCTACCCCATTTGGCTTAGGGCTGCATCGTGT 54704
      ::||| ::||| ::||| ::|||
403 .....Val.SerSerylProAla 408
      ::||| ::||| ::||| ::|||
54703 TCTAAGCTCAGACCTGCTTTATTATCACTTGGCTGTGAGTGTGGGCTTCGC 54654
      ::||| ::||| ::||| ::|||
409 ProSerProSerSerPheLeuProSerProSerProGlnProSerGlnSe 425
      ::||| ::||| ::||| ::|||
54653 CCATCTCCAGAGCTTCTCTCCCTAGCCCTTACACACAGCTTCTCAGAG 54604
      ::||| ::||| ::||| ::|||
425 rProValThrAlaArgThrProGlnAsnPheSerValProSerProGlyP 442
      ::||| ::||| ::||| ::|||
54603 CCCAGTGAAGAGAGCACCACCAAGAACTTACAGGCTTCTCCCTGAC 54554
      ::||| ::||| ::||| ::|||

```



```

442 rleuasnThrPro..... 446
|||||
54553 CTTTAACACCCCTGTAGTGGGCGTGGGTAGGACGCTTTACA 54504
446 ..... 446
54503 AGGTGCTTAATCCCTTCCTCCCAAGGGCTCAGATTGGTAAGGGCT 54454
446 ..... 446
54453 CATGGACCTCACAATCTCTGTACACCTGCTCCACAGGTTGACCAT 54404
447 .....ValasnProSerSerValMetSer 454
54403 CCATCTTACCAACCCGTCCTTATAGTGAACCCAGCTCTGTCAATGAGC 54354
445 ProAlaGlySerSerGlnAlaGlnGlnGlnTyrLeuAspLysLeuLys 471
54353 CCAGCTGGCTTAGCCAGGCTGAGAGCAGCAGTACTGACACAGCTGAA 54304
471 GlnLeuSerLysTyrLeuGlnProLeuArgArgMetLysLysLysLea 488
54303 GCATTTGTCACAGTACATCGAGCCCTCGCAGCATGATCAACAAGATCG 54254
488 sPLysAsnGlu..... 491
54253 ACAAGATATAAGTACGCTAGCCAAAGCAGATGGCGCCGACACCCAA 54204
492 .....AspArgLysLysAspLysLeu 498
54203 GGCCTGTTCTAGTGTCCCTTTCTTCCYAGACAGAAAAAGACTTAAG 54154
498 rLysMetLysSerLeuLeuAspLysLeuThrAspProSerLysArg.... 513
54153 TAAGATGAAGAGCTGCTGTGACATCTCCACGACCCCTCGAAGAGGTAG 54104
514 .....CysProLeuLysThrLeu..... 519
54103 TACTGGCCAGGGCTTTGGGCGAGGGCTGTGTGTGCGCACAGCTCACATAG 54054
520 .....GlnLysCys 522
54053 GGTCCCAAGACCTGTGACAGTATGATCAATCAAGGTCAACAGCG 54004
522 sGlnLysLeuLysLys.....LeuLysAsnAspMetArgCysProL 537
54003 CTCTGTGACCAAGCTGAGGCTTAGGTACTTTACAAAGAGGCTGTGACG 53954
537 euProHisArgProArgCysHisArgProAsnSerSerThrTyrAlaSer 553
53953 GGAAGCTTACCCCTCCCGGACCTT.....TCCAGCACCCACCTTTCA 53910
554 ArgSerTrpMetProSerTrpProThrSerAlaHisLeuSerSerThrL 570
53909 AGGAGTTGTGGT.....TTCACACACTT 53887
570 eProCysThrAlaHisSer 576
53886 CACTTGTGCTGTCAATAGC 53868

seq_name: gb_hlg:AC079044

seq documentation block:
LOCUS AC079044 228467 bp DNA HMG 10-AUG-2001
DEFINITION Mus musculus chromosome 16 clone rp23-5117, WORKING DRAFT SEQUENCE,
2 unordered pieces.
ACCESSION AC079044
VERSION AC079044.14 GI:15042857
KEYWORDS HMG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Mus.

```

```

REFERENCE 1 (bases 1 to 228467)
AUTHORS Yang,L., Hu,P., Wu,H., Prescott,A. and Roe,B.A.
TITLE Mus musculus BAC Clone rp23-5117
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 228467)
AUTHORS Yang,L., Hu,P., Wu,H., Prescott,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jul 31, 2001 this sequence version replaced gi:14861748.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 90075: contig of 90075 bp in length
* 90076 90175: gap of unknown length
* 90176 228467: contig of 136292 bp in length.
FEATURES
source
1..228467
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="16"
/clone="rp23-5117"
/clone_1b="RP23"
/location/Qualifiers
BASE COUNT 61424 a 52795 c 53888 g 60257 t 103 others
ORIGIN
alignment_scores:
Quality: 679.00 Length: 457
Ratio: 3.031 Gaps: 15
Percent Similarity: 49.015 Percent Identity: 41.357
alignment block:
US-09-668-119-3 x AC079044/rev ..
Align seg 1/1 to reverse of: AC079044 from: 1 to: 228467
276 ProProGlnProGlnProProValAlaGlnGlnProSerGlnLe 292
||| ||| ||||| :.....|||
51960 CCGAGGCTCCCTCTTATACACAGTGTGAGGAGGACGACATTACCAT 51911
292 uPro...ProGlnSerGlnThrGlnProLeuValSerGlnAlaGlnAlaL 308
: ||| :.....: ||||| :|||
51910 ATCTATACCAAGCCAGAAACCTGAATCTGTGTGTCTCCACCA... 51865
308 euProGlnGlnMetLeuTyrThrGlnPro.....ProLeuLysPhe 321
||| ||| ||| |||
51864 .....GGAGCCAGGCTACCTGACCTGCTTCTTCTGACCCCTTGATAC 51820
322 ValArgAlaProMetValAlaGlnGlnProProValAlaGlnProGlnValAl 338
||| ||| :.....: |||
51819 .....CATGTGTATGTGTGCTCCAGAGGTAAAGGCTGTGCTGAGTGAA 51776
338 nGlnGlnGlnThrAlaValAlaGlnThrAlaGlnAlaGlnMetValAlaP 355
: ||| :.....: |||
51775 AGAATAAGGGTGGATAAGATGGATGTGACAGCAGAGTGGGAGTTTGTAC 51726
355 roGlyVal.....GlnValSer 360
|||||
51725 AGGCTGTGTAGCATGAGCTGAGCTAAGTCTCTCCACTTCACAGTGAGC 51676
361 GlnSerSerLeuProMetLeuSerSerProSerProGlyGlnGlnValAl 377
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51675 CAAAGCAGCTTCACCATGCTGTCTCCACCGTACACGACGACGAGGTGCA 51626
377 nThrProGlnSerMetProProProGlnProSerProGlnProGly 394
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

51625 GACCCACAGTGTGATGACCCACACAGCGGCCACACACCTGGC. 51577
394 InProSerSerGlnProAsnSerSer..... 402
|||||
51576 .....TCACAGCCCAACTTAATGTCAGTAGGTGCTGCTGGATGAC 51535
402 ..... 402
51534 CCTTCCTTAGCCTCTTACCCCACTTTGGCCTTAGGGCTGCAATCGTGT 51485
403 .....Val.SerSerGlyProAla 408
|||||
51484 TCTAGGCTGACAGCCTGTTTATTCCTTACTTGGCTGAGCTCGGGCCCTGCC 51435
409 ProSerProSerSerPheLeuProSerProSerProGlnProSerGlnSe 425
|||||
51434 CCATCTCCAGCAGCTCTCTGCTAGCCCTTCACACAGCCCTTCACAGAG 51385
425 IProValThrAlaArgThrProGlnAsnPheSerValProSerProGlyP 442
|||||
51384 CCCAGTGAAGAGCAGCACCACCAAGAACTTCACGCTTCCCTCCCTGGAC 51335
442 IoleuAsnThrPro..... 446
|||||
51334 CTTTAAACACCCCTGTAAGTTGGGCGCTGGGTGGTAGGCAAGCTTTACA 51285
446 ..... 446
51284 AGGTGCTTAATCCCTTCCTCCACAGGCTCAGATTGGTTAAAGGCT 51235
446 ..... 446
51234 CATGGACCTCACAATCTCTGTGACACCTGCTCCACAGTTGAGCCAT 51185
447 .....ValAsnProSerSerValMetSer 454
|||||
51184 CCATCTAACCCACCTGTCCCTTAAGTGAACCCACGCTCTGTCATGAGC 51135
455 ProAlaGlySerSerGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 471
|||||
51134 CCAGTGGCTCTAGCCAGGCTGAGAGCAGCAGTACCTGGACAAGCTGAA 51085
471 GlnLeuSerLysTyrIleGlnProLeuArgMetIleAsnLysIleA 488
|||||
51084 GCAATTGTCAGATCATCGAGCCCTCGACGATGATCAACAAGATCG 51035
488 sPlysAsnGlu..... 491
|||||
51034 ACAGAAATGAAGGTAGGCTAGCCAAAGCAGAGATGGGCCGACACCCAA 50985
492 .....AspArgLysLysAspLeuSe 498
|||||
50984 GGCCTGTTCTTAGTGTCCCTTTCTTCTTACACAGAAAAAAGACTTAAG 50935
498 IlyMetLysSerLeuLeuAspIleLeuThrAspProSerLysArg... 513
|||||
50934 TAAATGAAGAGCCCTGTGACATCTCACCGACCCCTCGAAGAGGTGAG 50885
514 .....CysProLeuLysThrLeu..... 519
|||||
50884 TACTGGCCAGAGGCTTTGGGCAAGGCTGTGTGTGTCACACAGCTCACAATAG 50835
520 .....GlnLysCys 522
|||||
50834 GGTCCCAAGACTTGTGACAGTATGATTAATCAAAAGTCCAAAGCG 50785
522 sGluIleAlaLeuGluLys.....LeuLysAsnAspMetArgCysProL 537
|||||
50784 CTCTTGACCATGCTGAGGCTTAGGTACTTTACAGGTGAGGTGTCAG 50735
537 euProHisArgProArgCysHisArgProAsnSerSerThrTyrAlaSer 553
|||||
50734 GGAACCTTACCCCTCCGACACCTT.....TCACAGCACCACCTTTCA 50691

```

```

554 ArgSerTyrMetProSerTrpProThrSerAlaHisLeuSerSerThrIle 570
|||||
50690 AGGAGTTGTGCT.....TCACACTTT 50668
570 eProCysThrAlaHisSer 576
|||||
50667 CACTGTGCTGTCAATAGC 50649
seq_name: gp_hcrg:AL359452
seq_documentation_block:
LOCUS AL359452 64231 bp DNA HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-422P18, *** SEQUENCING IN
PROGRESS ***, 20 unordered pieces.
ACCESSION AL359452
VERSION AL359452.4 GI:9930964
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 64231)
REFERENCE
Plumb B.
Direct Submission
Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9864141.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba422p18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 52454 bases at least Q40
Consensus quality: 59753 bases at least Q30
Insert size: 6231; sum-of-contigs
Insert size: 190157; 8.2% error; agrose-fp
Quality coverage: 2.09x in Q20 bases; sum-of-contigs quality
coverage: 1.09x in Q20 bases; agrose-fp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2992: contig of 2992 bp in length
* 2993 3092: gap of 100 bp
* 3093 5136: contig of 2044 bp in length
* 5137 5236: gap of 100 bp
* 5237 7850: contig of 2614 bp in length
* 7851 7950: gap of 100 bp
* 7951 9976: contig of 2026 bp in length
* 9977 10076: gap of 100 bp
* 10077 12531: contig of 2455 bp in length
* 12532 12631: gap of 100 bp
* 12632 14688: contig of 2057 bp in length
* 14689 14788: gap of 100 bp
* 14789 18056: contig of 3268 bp in length
* 18057 18157: gap of 100 bp
* 18157 21784: contig of 3628 bp in length
* 21785 21884: gap of 100 bp
* 21885 24801: contig of 2917 bp in length

```


Fri Mar 1 09:12:02 2002

us-09-668-119-3.rge

Page 24

1634 GGAGATTCTCTGCAATCCTACACAGCGCGTGGCGCTGGAGACTTACTTAA 1683
521 yscysgluilealaleugluLysLeu 529
||||| |||||||
1684 AGTGTGAAAAAGCGCTGTGAGAAGATG 1709

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 19:26:58 ; Search time 26.5 Seconds
(without alignments)
1618.432 Million cell updates/sec

Title: US-09-668-119-3

Perfect score: 3010
Sequence: 1 MKRAGVAHSSKDMESHVF.....WPSAHSLSITPTAFSPQ 579

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101:*

- 1: /SIDS2/gcgcdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDS2/gcgcdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDS2/gcgcdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDS2/gcgcdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDS2/gcgcdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDS2/gcgcdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDS2/gcgcdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDS2/gcgcdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDS2/gcgcdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDS2/gcgcdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDS2/gcgcdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDS2/gcgcdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDS2/gcgcdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDS2/gcgcdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDS2/gcgcdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDS2/gcgcdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDS2/gcgcdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDS2/gcgcdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDS2/gcgcdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDS2/gcgcdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDS2/gcgcdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SIDS2/gcgcdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2746	91.2	748	22	AA040299
2	2583.5	85.8	780	22	AA042085
3	390	13.0	79	22	AA04262
4	390	13.0	79	22	AA042672
5	390	13.0	79	22	AA042672
6	382.5	12.7	1004	19	AA037856
7	382.5	12.7	1004	19	AA037856
8	376.5	12.5	149	22	AA072673
9	350.5	11.6	155	22	AA069612
10	345	11.5	145	22	AA069612
11	328.5	10.9	121	22	AA069609

12	322	10.7	2074	21	AA054319	Amino acid sequenc
13	312.5	10.4	3190	22	AA084634	Amino acid sequenc
14	312	10.4	2063	22	AA040064	Human polypeptide
15	311.5	10.3	389	19	AA062647	Mature durum wheat
16	310	10.3	2023	21	AA154320	Amino acid sequenc
17	304.5	10.1	2783	13	AA023962	AFP-1 (Ala 2460 Va
18	304.5	10.1	2783	13	AA023963	Human ZC3 protein.
19	296	9.8	1336	20	AA055933	Huntingtin accumu
20	295	9.8	113	22	AA069615	Human additional s
21	292.5	9.7	903	18	AA037389	Human additional s
22	292.5	9.7	903	18	AA037381	Kapost. s sarcoma-a
23	288	9.6	1162	21	AA058255	HHV8 ORF 73 protei
24	288	9.6	1162	21	AA058500	Amino acid sequenc
25	288	9.6	1162	22	AA062331	CREB binding prote
26	283	9.4	2441	16	AA079054	Cellular transcrip
27	283	9.4	2441	19	AA040058	Mouse nuclear CREB
28	283	9.4	2441	21	AA094252	Huntingtin accumu
29	282	9.4	123	22	AA069611	Mouse neural Mena+
30	281	9.3	783	19	AA037151	Mouse neural Mena+
31	280	9.3	787	19	AA037152	Mouse neural Mena+
32	280	9.3	802	19	AA037153	Human polypeptide
33	277.5	9.2	388	22	AA041453	Human polypeptide
34	277.5	9.2	576	22	AA039657	Peptide #124 encod
35	274.5	9.1	223	22	AA013630	Peptide #127 encod
36	274.5	9.1	223	22	AA026090	Peptide #122 encod
37	274.5	9.1	223	22	AA014440	Peptide #172 encod
38	272	9.0	170	22	AA021738	Peptide #12090 enc
39	272	9.0	170	22	AA038053	Shrimp white spot
40	271.5	9.0	2293	22	AA085028	Human secreted pro
41	271	9.0	55	21	AA027648	Human secreted pro
42	271	9.0	55	21	AA027901	Human secreted pro
43	270.5	9.0	395	17	AA030474	Mouse SVY-related
44	264	8.8	89	22	AA069608	Huntingtin accumu
45	260.5	8.7	420	21	AA053025	Human secreted pro

ALIGNMENTS

RESULT 1	
AA040299	standard; Protein: 748 AA.
AA040299	
22-OCT-2001	(first entry)
Human polypeptide SHQ ID NO 3444.	
Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; chromolytic; drug screening; arthritis; inflammation; leukaemia.	
Human sapiens.	
WO200153312-A1.	
26-JUL-2001.	
26-DEC-2000;	2000WO-US34263.
21-JAN-2000;	2000US-0488725.
25-APR-2000;	2000US-0552317.
09-JUL-2000;	2000US-0596042.
19-JUL-2000;	2000US-0620312.
03-AUG-2000;	2000US-0653450.
14-SEP-2000;	2000US-0662191.
19-OCT-2000;	2000US-0693036.
29-NOV-2000;	2000US-0727344.

(HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 N-PSDB: AA159453.

XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 XX Example 6; SEQ ID NO 3444; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA38642-AA42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: the sequence data for this patent did not form part of the printed
 CC specification.

Sequence 748 AA:

Query Match 91.2%; Score 2746; DB 22; Length 748;

Best Local Similarity 91.0%; Pred. No. 2e-161;

Matches 544; Conservative 10; Mismatches 16; Indels 28; Gaps 4;

QY 1 MRRAGVAHSKSKDMESHVELKAKTRDEYLSLVARLILHFRDHNKSSQASVSDPNALLO 60
 DB 27 mrragvshskskdmeshvelkaktreylsylvarlilfrdhnksgasvsdpnmalq 86
 QY 61 SLTGGGAAGAGIGMPRRGSGSLGCMGSGAMGPMSSGPPGTSGMAPMSMAVYST 120
 DB 87 sltggaagagagigmprrgsgslgcmgsgamgpmssgppgtsgmaphsmavest 146
 QY 121 ATPOTLOLQOVALCOOQO 180
 DB 147 atpqtglqgvalcsgq 206
 QY 181 QOQLOQO 240
 DB 207 qgqlqgq 266
 QY 241 QPPIOQPPMOQPPPPSQALPQOLOMHTNHQPPPOPPQPPVAVONQPSQLPQSQTOP 300
 DB 267 qppioqppmoqppppsqalpqolomhtnhqppppoppqppvavonqpsqlppsqtop 326
 QY 301 LVSQAQALPGQMLYTTPPLKFWAAWVVOQPPVQOVQOQOTAVOTAGAAGVAVAGVOVS 360
 DB 327 lvsqaqalpgqmlytppplkfwaaawvvoqppvqovqoqotavotagaagvavagvovs 386
 QY 361 GSSLPLMSSPQOQVOTPOSMPPPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 420
 DB 387 gsslplmssspqoqvotposmppppqppqppqppqppqppqppqppqppqppqppqppqpp 446
 QY 421 QSSQSVTARTPQNFVSPSPGFLNTPVNSVSPAGSQAEEQOYLDKIKOLSKYITPL 480
 DB 447 qssqsvtartpqnfvspspgflntpvnsvspagsqaeeqoyldkikolskyitpl 506
 QY 481 RRINKITDKNDRKKSLSKMSKSLDLIDPSKRCPLKTKCKEILKELKNDKRCPLDHR 540
 DB 507 rrminkitdkndrkkslskmskslldlidskrcplktklqcelaleklnomavprrpp 566

QY 541 P-----RCHRPNSSTYASRSMWSPWTSAN---LSSTIPCT 573
 DB 567 pvpptkqgylcgpdlldavlanirpvnfhsly--rtfypam-tahgppltapvc 621

RESULT 2

ID AAM42085 standard; Protein; 780 AA.

XX AAM42085;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 7016.

XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.
 XX N-PSDB: AA161241.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

Example 2; SEQ ID NO 7016; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA38642-AA42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: the sequence data for this patent did not form part of the printed
 CC specification.

Sequence 780 AA:

Query Match 85.8%; Score 2583.5; DB 22; Length 780;
 Best Local Similarity 86.0%; Pred. No. 2.2e-151;
 Matches 515; Conservative 13; Mismatches 42; Indels 29; Gaps 5;

QY 1 MKKAGVAHSSKSDMEHSHFLKAKTRDEFLSLVARLLIHFRDHNKKSQASVSDPMNLQ 60
 |||
 DB 58 MKKAGVAHSSKSDMEHSHFLKAKTRDEFLSLVARLLIHFRDHNKKSQASVSDPMNLQ 117
 QY 61 SLTGPAAGAGTGMPPRPGSLGMSFGAMGPMSTLSGPPPGTSGMAPHNAVVT 120
 |||
 DB 118 SLTGPAAGAGTGMPPRPGSLGMSFGAMGPMSTLSGPPPGTSGMAPHNAVVT 177
 QY 121 ATRPTQLQLQY-ALDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 179
 |||
 DB 178 ATRPTQLQLQY-ALDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 237
 QY 180 QDQ 239
 |||
 DB 238 QDQ 297
 QY 240 AORPTQORPQORPQORPQORPQORPQORPQORPQORPQORPQORPQORPQ 299
 |||
 DB 298 AORPTQORPQORPQORPQORPQORPQORPQORPQORPQORPQORPQORPQ 357
 QY 300 PLVSGQALPGQMLYTOPPLKFRAPVVOOPVOPVOQOQOTAVOTAAQAVAVGV 359
 |||
 DB 358 PLVSGQALPGQMLYTOPPLKFRAPVVOOPVOPVOPVOPVOPVOPVOPVOPV 417
 QY 360 SSSSLMLSSSPGQOVOPQOSMPRPPQOPQOPQOPQOPQOPQOPQOPQOPQOP 419
 |||
 DB 418 SSSSLMLSSSPGQOVOPQOSMPRPPQOPQOPQOPQOPQOPQOPQOPQOPQOP 477
 QY 420 PQSOSPVARTPQNFVSPSPGLNTPVNPSSVMSFAGSSQAEQOYLDKLOLKTIEP 479
 |||
 DB 478 PQSOSPVARTPQNFVSPSPGLNTPVNPSSVMSFAGSSQAEQOYLDKLOLKTIEP 537
 QY 480 LRMINKTDKNEDEKDKSLDKSLDLITDPKRCPLTKCEIKLNDKNDREPLH 539
 |||
 DB 538 LRMINKTDKNEDEKDKSLDKSLDLITDPKRCPLTKCEIKLNDKNDREPLH 597
 QY 540 RP-----RCHRPNSSTYASRSMWSPMPTSAH--LSSTIPCT 573
 |||
 DB 598 RP-----RCHRPNSSTYASRSMWSPMPTSAH--LSSTIPCT 653

RESULT 3
 ID AAM14262 standard; Protein; 79 AA.
 AC AAM14262;
 DT 12-OCT-2001 (first entry)
 DE Peptide #696 encoded by probe for measuring cervical gene expression.
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 OS Homo sapiens.
 PN WO200157278-A2.
 PD 09-AUG-2001.
 PR 30-JAN-2001; 2001WO-US00670.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488901/53.
 DR Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -
 PT Claim 27; SEQ ID No 19088; 487bp; English.
 PS The present invention relates to human single exon nucleic acid probes
 CC (SNP: see A110068-A1128459). The present sequence is a peptide encoded
 CC by one such probe. The SEMP are derived from human HeLa cells. The SEMP
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 79 AA:

Query Match 13.0%; Score 390; DB 22; Length 79;
 Best Local Similarity 100.0%; Pred. No. 2.3e-17;
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 QLOQVALLQDQ 185
 |||
 DB 1 126 QLOQVALLQDQ 60
 QY 186 QDQ 204
 |||
 DB 61 QDQ 79

RESULT 4
 ID AAM26672 standard; Protein; 79 AA.
 AC AAM26672;
 DT 17-OCT-2001 (first entry)
 DE Peptide #709 encoded by probe for measuring placental gene expression.
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 OS Homo sapiens.
 PN WO200157272-A2.
 PD 09-AUG-2001.
 PR 30-JAN-2001; 2001WO-US00663.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;

identifying neoplastic tissue of a human, where over-expression of the hphl gene identifies the tissue as being neoplastic; (7) a method to aid in the diagnosis or prognosis of neoplasia, where a difference in the hphl gene, mRNA, or protein between a first and second tissues indicates neoplasia in the first tissue; (8) a method to aid in detecting a genetic predisposition to neoplasia; (9) a method of identifying a human chromosome 12; and (10) a method of inducing a cell to de-differentiate by contact with a hphl gene or expression product. The products can also be used to inhibit hphl expression to suppress neoplasia, dysplasia, or hyperplastic cell growth. They can be used to treat e.g. tumours, amygdalic hereditary ectodermal dysplasia, congenital alveolar dysplasia, epithelial dysplasia of the cervix, fibrous dysplasia of bone, or mammary dysplasia, endometrial, adrenal, breast, prostate, or thyroid hyperplasia or pseudo-epitheliomatous hyperplasia of the skin. Even in disorders in which hphl mutations are not implicated, down-regulation or inhibition of hphl expression can have therapeutic application. In these disorders, decreasing hphl expression can help to suppress tumours. Similarly, in tumours where hphl expression is not aberrant, effecting hphl down-regulation can suppress metastases.

Query Match	12.7%;	Score 382.5;	DB 19;	Length 1004;
Best Local Similarity	28.1%;	Pred. No. 1e-15;		
Matches 151;	Conservative 67;	Mismatches 218;	Indels 101;	Gaps 22

[illegible]

RESULT	7
AAW52830	
ID	AAW52830 standard; Protein; 1004 AA
XX	
AC	AAW52830;
XX	
DT	09-JUL-1998 (first entry)
XX	

XX		
XX	A	tumour suppressor gene called polyhomeotic 1 (pnh1).
KM		
KM	Tumour suppressor gene; polyhomeotic 1; hph1: human chromosome 12p13;	
KM	Identification; neoplastic tissue; cellular differentiation; diagnosis	
KW	neoplasia; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	W09807858-A1.	
XX		
PD	26-FEB-1998.	
XX		
PE	22-AUG-1997;	97WO-US14866.
XX		
PR	06-FEB-1997;	97US-0036339.
PR	23-AUG-1996;	96US-0024349.
PR	04-DEC-1996;	96US-0031569.
XX		
PA	(CHIR)	CHIRON CORP.
XX		
PI	Randazzo F;	
XX		
DR	WPI: 1998-169162/15.	
DR	N-PSDB: AAV21060.	
XX		
PT	Isolated human poly:homeotic 1 oncogene - used to develop products	
PT	for diagnosis and therapy of proliferative and developmental	
PT	disorders, e.g. neoplasia, dysplasia or hyperplasia	
XX		
PS	Claim 1; Pages 28-30; 40pp; English.	

CC The present sequence encodes a novel human tumour suppressor gene
CC termed polyhemicin 1 (hph1). The hph1 gene maps to human chromosome
CC 12p13, a region which is frequently lost in non-small cell lung cancer
CC and breast cancer. A method of identifying neoplastic tissue of a
CC human comprised comparing the expression of a hph1 gene in a tissue of
CC a human suspected of being neoplastic with the expression of a hph1 gene
CC in a tissue of the human which is normal. Under-expression of the hph1
CC gene identifies the subject as having neoplastic tissue. The hph1
CC oncogene functions to suppress neoplasia and dysplastic or hyperplastic
CC cell growth as well as to induce cellular differentiation. The cDNA,
CC protein and vectors can be used as diagnostic and therapeutic tools for
CC proliferative and developmental disorders and to identify a p13 region
CC of a human chromosome 12. They can be used for the detection, diagnosis
CC or prognosis of neoplasia or for detecting a genetic predisposition to
CC neoplasia. They can also be used to treat tumours.

[illegible]

OY 342 TAVCTAAQAAWAPVGVVSSSLPMLSSPSPGQVOTPOSMPPPPSPGPPSSQPN 401
 Db 530 vtirpqlagqgajlajlaavp-----srgmptvq-sgqhlaspss 576
 OY 402 NVSSGPAPSPSPSPSPSPSQ-----SPVTRTRPQNFVSPSPGPNTPVN 448
 Db 577 qapgalgcptlla2gmrlapvgtahvkgatssvvaqvpafymgs---vhlpgk 633
 OY 449 PSSVSPAGSSQAEEOQVLDKLTOLSKYIEPLRMINKIDKEDRKKDL-SKMSLIDL 507
 Db 634 pqlt---avtkkadseerdvstclgmtpakaspvaeapkmeksslgkkaesvanv- 689
 OY 508 TDPKRCPLKTLQK-ETALEKIKNDKRCPLPBRPCHRPNSSTYASBSMPSMPTSA 564
 Db 690 ---nanpssel-----lpsvvp-----ppllamvstrgmgsdkpqa 727

RESULT 8

AAB72673
 ID AAB72673 standard; protein; 149 AA.

AC AAB72673;

DT 09-MAY-2001 (first entry)

DE Polyglutamine tract #2.

KW Animal model: polyglutamine tract; neurodegenerative disorder: HDJ1;
 KW heat shock protein 40; tetratricopeptide repeat protein 2; TPR2; stroke;
 KW myeloid leukaemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
 KW Parkinson's disease; CJD; BSE; Huntington's disease; head trauma; cancer.
 OS Drosophila sp.

PN WO200112238-A1.

PD 22-FEB-2001.

PF 14-AUG-2000; 2000WO-0522496.

PR 12-AUG-1999; 99US-0148933.

PR 12-AUG-1999; 99US-0148934.

PR 18-JAN-2000; 2000US-0177047.

PR 19-MAY-2000; 2000US-0205720.

(CALY) CALIFORNIA INST OF TECHNOLOGY.

PI Benzer S, Kazemi-Esfarjani P;

DR WPI; 2001-147537/15

PT Identifying genes of other compounds that modulate polyglutamine
 PT toxicity, useful for treating Alzheimer's disease, Parkinson's disease
 PT and Creutzfeldt-Jakob disease -

PS Example 2; Fig 1B; 775pp; English.

XX The present invention describes a method of screening for genes which
 CC modulate polyglutamine toxicity using animal models with polyglutamine
 CC sequences that cause toxicity in the animal. The model is preferably
 CC Drosophila, and toxic polyglutamine sequences include the human and
 CC Drosophila heat shock protein 40/HDJ1, tetratricopeptide repeat protein 2
 CC (TPR2) and myeloid leukaemia factor 1 (MLF) genes. The model is useful
 CC for identifying treatments for neurodegenerative and proliferative
 CC disorders, including Alzheimer's disease, Parkinson's disease,
 CC Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
 CC (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar
 CC ataxias, dentatorubropallidoluysian atrophy, Kennedy's disease, stroke,
 CC head trauma and cancer.

SQ Sequence 149 AA;

Query Match

Best Local Similarity 12.5%; Score 376.5; DB 22; Length 149;

Matches 91; Conservative 5; Mismatches 46; Indels 21; Gaps 2;

OY 99 LSGPPPTGSGMPSHMAVAVSTATPOTOLQLOVALQOQOQOQOQOQOQOQOQ 158
 Db 1 mgspp-----stpqgqgqgqgqgqgqgqgqgqgqgqgqgqgqgqgq 44

OY 159 QQQQFQAQSSAMQGFQAVVQOQOQLOQOQOQOQOQOQOQOQOQOQOQOQ 218
 Db 45 qqqqgq 99

OY 219 QLOQ 261
 Db 100 qqqgq 142

RESULT 9

AAB69612
 ID AAB69612 standard; protein; 155 AA.

AC AAB69612;

DT 30-APR-2001 (first entry)

DE Huntingtin accumulation inhibitor peptide HD-Q104-Myc-HIS6.

KW Neurological disorder; Huntington's disease; Alzheimer's disease;
 KW Parkinson's disease; prion disease; frontotemporal dementia;
 KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
 KW dentatorubral-pallidoluysian atrophy; spinocerebellar ataxia type 1;
 KW SCA2; SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.

OS Synthetic.

PN WO200106989-A2.

PD 01-FEB-2001.

PF 24-JUL-2000; 2000WO-US20131.

PR 27-JUL-1999; 99US-0146047.

PR 21-JUL-2000; 2000US-0620955.

(HUST/) HUSTON J S.

PA (MESS/) MESSER A.

PA (LECER/) LECERF J.

PI Huston JS, Messer A, Lecerf J;

DR WPI; 2001-182700/18.

PT Inhibiting intracellular polypeptide accumulation, useful for treating
 PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
 PT the polypeptide with a specific intrabody -

PS Disclosure: Page 99; 108pp; English.

XX The present invention describes a method for inhibiting the formation of
 CC aggregates of certain proteins, involving contacting the protein with a
 CC binding molecule known as an intrabody. Proteins to be bound include
 CC those associated with neurological disorders, and so the method can be
 CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
 CC Huntington's diseases, prion disease, frontotemporal dementia,
 CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
 CC dentatorubral-pallidoluysian atrophy, spinocerebellar ataxia type 1
 CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7.

SQ Sequence 155 AA;

Query Match

11.6%; Score 350.5; DB 22; Length 155;

'Fri Mar 1 09:12:01 2002

us-09-668-119-3.rag

1
2
3
4

—
—

Fri Mar 1 09:12:01 2002

us-09-668-119-3.ra1

Page 10

alignment scores:
 Quality: 2988.00 Length: 580
 Ratio: 5.161 Gaps: 1
 Percent Similarity: 99.828 Percent Identity: 99.483

alignment block:
 US-09-668-119-3 x AA159455

Align seg 1/1 to: AA159455 from: 1 to: 2556

```

1 MetArgValAlaGlyValAlaHisSerLysSerLysAspMetGluSe 17
338 ATGAGGAAAGCTGGTGTGGCACACAGTAATTCACCAAGGATATGGAGAG 387
17 rHisValPheLeuLysAlaLysThrArgAspGluThrLeuSerLeuVala 34
388 CCATGTTTCTCGTAAAGCCCAAGACCCGGAGCAATACCTTCTCTGCGTGG 437
34 laArgLeuIleIleHisPheArgAspIleHisAsnLysLysSerGlnAla 50
438 CCAGGCTCATATATCCATTTCGAGACATTCAACCAAGAAATCTCAAGCT 487
51 SerValSerAspProGluAsnAlaLeuGlnSerLeuThrGlyProAla 67
488 TCCGTCAGATGATCCATTCATGCACTCCAGACCTGAGCTGGGACCTGC 537
67 alaArgValAlaAlaGlyValIleGlyMetProProArgGlyProGlyInserL 84
538 TGGCGGAGACCCGCTGGATTGGCATGCTCTCTGGGGCCCGGGAGACGTC 587
84 euGlyGlyMetGlyAsnThrPheGlyAlaMetGlyGlnProMetSerLeuSer 100
588 TGGGGGATGGGATGACCTGTGGTGGACAGGACAGCCAAATGCTCTCTCA 637
101 GlyGlnProProGlyThrSerGlyMetAlaProHisSerMetAlaVala 117
638 GGGCAGCGCCCTCCCTGGACCTCGGGGATGGCCCTCACACAGATGGCTGT 687
117 lValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAlaL 134
688 CGTGTCTACAGGCAACCTCCACAGACCCAGCTCCAGCAGAGGTGGCGC 737
134 euGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 150
738 TGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 787
151 LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 167
788 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 837
167 nSerAlaMetGlnGlnGlnPheGlnAlaValaGlnGlnGlnGlnGln 184
838 GAGTCGATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 887
184 euGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 200
888 TCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 937
201 AsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 217
938 AATTCAGCAACAGATATACAGCAGCAGCAACACACCTGACGCAATAGACA 987
217 nLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 234
988 GCTGACACTCCCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1037
234 lngGlnGlnAlaLeuGlnAlaGlnProProIleGlnGlnProProMetGln 250
1038 AGCAGAGGCTTTGCAAGGCCACCAACCAATTGACGACGACGACGATGCA 1087
251 GlnProGlnProProProSerGlnAlaLeuProGlnGlnGlnGlnGlnGln 267
1088 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1137
267 tHisHisThrGlnHisHisGlnProProProGlnProGlnProGlnPro 284
1138 GCATCAGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1187
284 alaAlaGlnAsnGlnProSerGlnLeuProProGlnSerGlnThrGlnPro 300
1188 TTGCTCAGAACCAACCATCAACTCCCGCACAGCTCCGACAGACCCAGCT 1237
301 LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuThrGlnPro 317
1238 TTGGTGTCAACAGCGCGCAGGCTCTCCCTGGCAAAATGTTGATACCAACC 1287
317 oProLeuLysPheValArgAlaProMetValAlaGlnGlnProProValG 334
1288 ACCACTGAATTTGTCCGAGCTCCGATGGTGGTGCAGCAGCAGCAGCTGC 1337
334 lProGlnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 350
1338 AGCCCGAGGTGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1387
351 GlnMetValAlaProGlyValGlnValSerGlnSerSerLeuProMetLe 367
1388 CAGATGTGTGCTCCGAGTCCAGGTCAGGTCAGCAGCAGCAGCAGCTCC 1437
367 uSerSerProSerProGlyGlnGlnValGlnThrProGlnSerMetPro 384
1438 GTCTCTCCGCTCACCGGGCAGCAGGTCAGACCCCGCAGTGTGATGCC 1487
384 roProProGlnProSerProGlnProGlyGlnProSerSerGlnProAsn 400
1488 CTCGCCCCAGCGCTCCCGCAGCCCGGACAGCCAGCTCACAGCCCAAC 1537
401 SerAsnValSerSerGlyProAlaProSerProSerSerPheLeuProSe 417
1538 TCCAAACGTAGCTGTGGCTCCGCTCCATCTCCAGTAGTCTCTCCAG 1587
417 rProSerProGlnProSerGlnSerProValThrAlaArgThrProGln 434
1588 CCCCTACCGCAGCCCTCCACAGCCAGTGCAGGCGGAGCCACACAGA 1637
434 snPheSerValProSerProGlyProLeuAsnThrProValAsnProSer 450
1638 ACTTCAGTGTCCCTACCTGAGACCTTTAAACACACCTGTGAACCCAGC 1687
451 SerValMetSerProAlaGlySerSerGlnAlaGlnGlnGlnGlnGln 467
1688 TCTGTATATAGCCAGCTGGCTCCAGCCAGGCTGAGGAGCAGCACTACT 1737
467 uAspLysLeuLysGlnLeuSerLysThrLeuGlnProLeuArgArgMet 484
1738 GGACAAAGCTGGAAGCAGCTGTCCGAATACATCAGACCCCTGCGCCGCA 1787
484 leAsnLysIleAspLysAsnGlnuAspArgLysLysAspLeuSerLysMet 500
1788 TCAACAAGATCGACAAGAAGCAGACAGCAAAAGAGACTGTGATAGATG 1837
501 LysSerLeuLeuAspIleLeuThrAspProSerLysArgCysProLeuLys 517
1838 AAGAGCCTTCTGAGCATTTGACACAGCCCTCGAAGCGGTGTCCCTGAA 1887
517 sThrLeuGlnLysCysGlnIleAlaLeuGlnLysLeuLysAsnAspMet 533
1888 GACCTTGCAAAAGTGTGATGCTGCTGGAAGAACTCAAGATGACATGG 1937
534 ArgCysProLeuProHisArgProArgCysHisArgProAsnSerThr 550
1938 CGGTCCCACTCCCGCAGCCCGGCGGACCAACCAACCAACCAACCAAC 1987
550 rTyrAlaSerArgSerThrMetProSerThrProThrSerAlaHisLeuS 567

```

1988 CTATGCCAGCCGCTCCTGATGCCGTCCTGCGCCACATCCGCTCACCCTG 2037

567 eiserThrIleProCysThrAlaHisSerPheGlnPro 579

2038 CTTCACACATTCCTCCGTACCGCACATTCGTTCCAGCCA 2075

seq_name: /SID57/gcgdata/geneseq/geneseqn/MA2001.DAT:AA161241

seq_documentation_block:

ID AA161241 standard; CDNA; 3438 BP.

AA161241;

22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 5230.

Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukemia; ss.

Homo sapiens.

W020015312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QH, Zhou P, Goodrich R, Drmanac RT;

WPI; 2001-442253/47.

P-PSDB; AAM42085.

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

Claim 1; SEQ ID NO 5230; 10078bp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAM42213) with nocotropic,

immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide

of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

utilization of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukemias and

C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification.

Sequence 3438 BP; 771 A; 1143 C; 932 G; 592 T; 0 other;

alignment_scores:

Quality: 2886.00 Length: 587

Ratio: 4.993 Gaps: 8

Percent Similarity: 98.467 Percent Identity: 97.956

alignment_block:

US-09-668-119-3 x AA161241 ..

Align seg 1/1 to: AA161241 from: 1 to: 3438

1 MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetGlu 17

338 ATGAGGAAGCTGTGTGGCACAGTAATCCAGCAAGATATGAGAG 387

17 HisValPheLeuLysAlaLysThrArgAspLysLeuSerLeuVal 34

388 CCATGTTTCCGAGGCCAGACCCGGGAGCAATACCTTCTCTGAG 437

34 LArgLeuIleIleHisPheArgAspIleHisAsnLysSerGlnAla 50

438 CCAAGCTCATTTATCCATTTTCAGACATTCATACAGAAATCTAGCT 487

51 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyProAl 67

488 TCCGTCAGATCCTATGATGACACTCCAGAGCTGAGCGGACCTGC 537

67 AlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSer 84

538 TGGGGAGCCCGTGGAAATGGCATGCTCTCGGGGCCCGGACAGTCT 587

84 euGlyIleMetGlySerPheGlyAlaMetGlyGlnProMetSerLeuSer 100

588 TGGGGGAGATGGGTAGCTGTGGTGCATGGAGCAGCAATGCTCTCA 637

101 GlyGlnProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 117

638 GGGCAGCCGCTCTCGGACCTCGGGGATGGCCCTCAGCAGATGGCTGT 687

117 ValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAla 133

688 CGTGTCTACGGCACTCCACAGACCAGCTGCAGCTCCAGAGTGGCG 737

134 LeuGlnGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnAla 150

738 CTGCAGCAGCAGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCG 787

150 LaleuGlnGlnGlnGlnGlnGlnGlnGlnGlnPheGlnAlaG 166

788 CGCTACAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAGCTCCAGCTCA 837

166 ngInSerAlaMetGlnGlnGlnPheGlnAlaValValGlnGlnG 182

838 GCAGAGTGCATGCAGCAGCAGCAGCTTCCAGCAAGTGTGAGAGCA 887

182 ngInLeuGlnGlnGlnGlnGlnGlnGlnGlnHisLeuHis 198

888 GCAGCTTCCAGCAGCAGCAGCAGCAGCAGCAGCATTAATTAATGCA 937

199 HisGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnArg 215

938 CATCAAAATGCAACAGATACAGCAGCAGCAGCAGCAGCTGAGGAAT 987

215 eAlaGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 232

988 AGCAGACAGCTGAGTCCACACAGCAGCAGCAGCAGCAGCAGCAG 1037

232 lngGlnGlnGlnGlnAlaLeuGlnAlaGlnProProIleGlnProPro 248

1038 AGCAGCAGCAGCAGCAGCTTTCAGGCCAGCCAGCAGCAATTCAGCAGCCG 1087

249 MetGlnGlnProGlnProProSerGlnAlaLeuProGlnGlnLeuG 265

```

|||||
1088 ATGCAGCAGCCACAGCTCTCCGCGCTCCAGAGCTCTGCCCGCAGCAGCTGCA 1137
|||||
265 NGIMETHSHSHSTHrGlnHSHSHSGLnProProProGlnInProGlnInP 282
|||||
1138 GCAGTGTGTGACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1187
|||||
292 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 298
|||||
1188 CTCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1237
|||||
299 GlnProLeuValSer3InlAglInlAleuProGlyGlnMetLeuTyrTh 315
|||||
1238 CAGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1287
|||||
315 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1337
|||||
1288 CCAACCCAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1337
|||||
332 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1348
|||||
1338 CAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1387
|||||
349 ATATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 365
|||||
1388 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1437
|||||
365 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 382
|||||
1438 CATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1487
|||||
382 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 398
|||||
1488 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1537
|||||
399 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 415
|||||
1538 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1587
|||||
415 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1637
|||||
1588 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1687
|||||
432 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1737
|||||
1638 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1787
|||||
448 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1837
|||||
1688 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1887
|||||
465 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1937
|||||
1738 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1987
|||||
482 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2037
|||||
1788 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2082
|||||
498 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2137
|||||
1838 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2187
|||||
515 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2237
|||||
1888 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2282
|||||
532 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2337
|||||
1938 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2382
|||||
548 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2437
|||||

```

```

1988 GCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2037
565 H1sLeuSerSerThrlleProCysThrlAhlIsSerPheGlnPro 579
|||||
2038 CAGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2082
seq_name: /gicdata/geneseq/geneseq/NA2000.DAT: AAC59039
seq_documentation_block:
ID AAC59039 standard; cDNA: 1388 BP.
XX
AC AAC59039;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein coding sequence SEQ ID NO: 60.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200055175-A1.
XX
PD 21-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06049.
XX
PR 12-MAR-1999; 99US-0124144.
PR 11-JUN-1999; 99US-0138574.
PR 03-DEC-1999; 99US-0168667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI: 2000-638175/61.
XX
DX P-PSDB: AAB27609.
XX
PM Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 1; Page 372; 428pp; English.
XX
CC The invention relates to the isolation of genes AA58990-A59039 encoding
CC 50 human secreted proteins AAB27560-B27609. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (SEQID1) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune;
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 1388 BP; 364 A; 378 C; 343 G; 297 T; 6 other;

```

alignment_scores:
 Quality: 1597.00 Length: 319
 Ratio: 5.006 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.746

alignment_block:
US-09-668-119-3 x AAC59039 ..

Align seg 1/1 to: AAC59039 from: 1 to: 1388

```

1 MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetLys 17
  |||
99 ATGAGGAAGCTGTGTGGACA.AGTAAATCCAGCAAGATATGGAGAG 147
17 rHisValPheLeuLysAlaLysThrArgAspGluTyrLeuSerLeuValA 34
  |||
148 CCATGTTTCTCTGAGGCCAGACGCCGGAGCATACCTTCTCTGCTGG 197
34 laArgLeuIleIleHisPheArgAspIleHisAnLysLysSerGlnAla 50
  |||
198 CCAGGCTCATTCATTTCCAGACATTCATCAACAAGAAATCTCAAGCT 247
51 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyGlyProAl 67
  |||
248 TCCGTAGATATCTATGAATGCATCCAGAGCTGATGCGGACCTGC 297
67 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSerL 84
  |||
298 TCGGGAGAGCGGCTGCAATTGGCATGCTCTCGGGGCCGGGAGACGTC 347
84 euGlyLysMetGlySerPheGlyAlaMetGlyLysProMetSerLeuSer 100
  |||
348 TGGCGCGGATGGGTAGCCTGTGGCATGGGACAGCCAAATGCTCTCTCA 397
100 GlyGlnProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 117
  |||
398 GGGGAGCGGCTCTGGGACCTCGGGATGGCCCTCAGCAGATGGGTGT 447
117 lValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnAlaValAla 134
  |||
448 CGTGTCTACGGCACTCCAGACAGCCAGCTGACGCTCAGCAGGTGGCG 497
134 euGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnGlnGlnAla 150
  |||
498 TGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCGCG 547
151 LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 167
  |||
548 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCTCA 597
167 nSerAlaMetGlnGlnGlnPheGlnAlaValAlaGlnGlnGlnGln 184
  |||
598 GAGTGCATGACAGCAGCAGTCCAGACAGATAGTCACAGCAGCAGCAGC 647
184 euGlnGlnGlnGlnGlnGlnGlnHisLeuIleLysLeuHisHisGln 200
  |||
648 TCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTCAATTAATTCATCA 697
201 AsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 217
  |||
698 AATACGACACAGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 747
217 nLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 234
  |||
748 GCTGAGCTTCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 797
234 GlnGlnAlaLeuGlnAlaGlnProProIleGlnGlnGlnGlnGln 250
  |||
798 AGCAGCAGCAGCTTTCAGCGCCAGCAGCAGCAGCAGCAGCAGCAGC 847
251 GlnProGlnProProSerGlnAlaLeuProGlnGlnGlnGlnGln 267
  |||
848 CAGCCACAGCTCCGCTCCAGGCTGCTCCACAGCAGCAGCAGCAGAT 897
267 lHisHisThrGlnHisHisGlnProProGlnGlnGlnGlnGlnGln 284
  |||

```

```

898 GCATCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 947
284 aAlaGlnAsnGlnProSerGlnLeuProProGlnSerGlnThrGlnPro 300
  |||
948 TTGCTCAGACACACATCACACTCCGCCAGCTGCGACAGCAGCAGCAGCT 997
301 LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuTyrThrGlnPr 317
  |||
998 TTGGTGTACAGAGCGCCAGAGCTCTCCCTGACAAATGTGTATACCAACC 1047
317 oProLeu 319
  |||
1048 ACCAAAT 1054
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF65986
seq_documentation_block:
ID   AAF65986 standard; cDNA; 386 BP.
XX
XX   AAF65986;
AC
XX
XX   09-APR-2001 (first entry)
DT
XX
XX   Novel human polynucleotide, SEQ ID NO: 1742.
DE
XX
XX   Human; cytosstatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
XX   Homo sapiens.
XX
XX   MO200102568-A2.
XX
XX   11-JAN-2001.
XX
XX   30-JUN-2000; 2000MO-US18374.
XX
XX   02-JUL-1999; 99US-0142310.
XX
XX   02-JUL-1999; 99US-0142311.
XX
XX   (CHIR ) CHIRON CORP.
XX
XX   (HRSE-) HRSEQ INC.
XX
XX   Williams LT, Escobedo J, Innis MA, Garcia PD, Klingler J, Kassam A;
PI   Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI   Ckenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI   Kita D, Garcia V, Jones LW, Strache-Grain B;
XX
XX   WPI; 2001-091805/10.
XX
XX   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
XX   Claim 9; Page 793; 1046pp; English.
XX
XX   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
XX   Sequence 386 BP; 50 A; 88 C; 138 G; 110 T; 0 other;

```



```

XX Probe: microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.
XX Homo sapiens.
XX W0200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00663.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.
XX Claim 25; SEQ ID NO 719; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 436 BP; 105 A; 127 C; 117 G; 87 T; 0 other;
XX
alignment_scores:
XX Quality: 413.50 Length: 111
XX Ratio: 4.307 Gaps: 1
XX Percent Similarity: 86.486 Percent Identity: 78.378
XX
alignment_block:
XX US-09-668-119-3 x AA132033 ..
XX
Align seg 1/1 to: AA132033 from: 1 to: 436
XX
95 GlnPrometSerleuSerGlyGlnProProGlyThrSerGlyMetAl 111
123 CAGCCGTCGAGCGTTCTGGGAACTGCTGCTGAGTGGAGTGC.... 167
111 apProHisSerMetAlaValAlaSerThrAlaThrProGlyThrGlnLeu 128
168 .....GACTCTGCTCTTTCTCAGCCAGCTGC 195
128 InLeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGln 144
196 AGCTTCAGCAGGTGGCGCTGCAGCAGCAGCAGCAGCAGTTCAG 245
145 GlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGln 161
246 CAGCAGCAGCAGCGCGCTACAGCAGCAGCAGCAGCAGCAGCA 295
296 GCAGTTCAGGCTCAGCAGAGTGCATGACAGCAGCAGTTCAG 345
178 aGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeu 194
346 TGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCA 395

```

```

195 IleuLeuHisHisGlnAsnGlnGlnGlnLeu 205
396 ATTAATTCATCATCAATAATCAGCAGCAGTCA 428
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT.AA100701
seq_documentation_block:
ID AA100701 standard; DNA: 436 BP.
XX AA100701:
XX 09-OCT-2001 (first entry)
XX Probe #692 used to measure gene expression in human breast sample.
XX Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX Homo sapiens.
XX W0200157270-A2.
XX 09-AUG-2001.
XX 29-JAN-2001; 2001WO-US00661.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast.
XX Claim 25; SEQ ID NO 692; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 436 BP; 105 A; 127 C; 117 G; 87 T; 0 other;
XX
alignment_scores:
XX Quality: 413.50 Length: 111
XX Ratio: 4.307 Gaps: 1
XX Percent Similarity: 86.486 Percent Identity: 78.378
XX
alignment_block:
XX US-09-668-119-3 x AA100701 ..
XX
Align seg 1/1 to: AA100701 from: 1 to: 436

```

95 GlnPrometSerLeuSerGlyGlnProProGlyThrSerGlyMetAl 111
 123 CAGCCCTCAGACGTTCTGGGAATCTGCGCTGAGTGCAGTGC... 167
 111 aProHisSerMetAlValValSerThrAlaThrProGlnThrGlnLeu 128
 168GACTGAGTCTTTCTTCAGCCCAAGCTGC 195
 128 LInLeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGln 144
 196 AGCTCCAGCAGAGTGGCTGCTGCAAGCAGCAGCAACAGCAGTTCAG 245
 145 GlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGln 161
 246 CAGCAGCAGCAGCAGCGCGCTCAGCAGCAGCAGCAGCAGCA 295
 161 nGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnPheGlnAlaVal 178
 296 GCAGTTCAGGCTCAGCAGAGTGCATGCAAGCAGCAGTTCAGCAGTAG 345
 178 aGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 194
 346 TGCAGCAGCAGCAGCAGCGCTCAGCAGCAGCAGCAGCAGCAGCAGTCTA 395
 195 IleLeuHisHisGlnHisGlnGlnGlnGlnGlnGlnGln 205
 396 ATTAATTCATCATCAAAATCAGCAGAGGTA 428

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI20017

seq_documentation_block:
 ID AAI20017 standard; LNA; 316 BP.

AAI20017;

12-OCT-2001 (first entry)

Probe #9950 for gene expression analysis in human cervical cell sample.

Probe; human; microarray; gene expression; cervical epithelial cell;

cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID NO 9950; 487bp; English.

CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 316 BP; 81 A; 104 C; 87 G; 44 T; 0 other;

alignment_scores:
 Quality: 393.00 Length: 80
 Ratio: 4.912 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.750

alignment_block:
 US-09-668-119-3 x AAI20017

Align seg 1/1 to: AAI20017 from: 1 to: 316

126 GlnLeuGlnLeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGln 142
 3 CAGCTCAGCTCCAGCAGAGTGGCTGCAAGCAGCAGCAGCAACAGCAGCA 52
 142 nPheGlnGlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGln 159
 53 GTTCCAGCAGCAGCAGCGCGCTCAGCAGCAGCAGCAGCAGCAGCAGC 102
 159 GlnGlnGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnPheGln 175
 103 AACAGCAGCAGTTCAGGCTCAGCAGAGTGCATGCAAGCAGCAGTTCAG 152
 176 AlaValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 192
 153 GCAGTAGTCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCA 202
 192 nHisLeuIleLeuHisHisGlnHisGlnGlnGlnGlnGlnGln 205
 203 GCATCTAATTAATTCATCATCAAAATCAGCAGCAGGTA 242

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI5215

seq_documentation_block:
 ID AAI5215 standard; DNA; 316 BP.

AAI5215;

17-OCT-2001 (first entry)

Probe #13901 used to measure gene expression in human placenta sample.

Probe; microarray; human; placenta; antenatal diagnosis;

genetic disorder; ss.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;


```

89  erPheGlyAlaMetGlyGlnProMetSerLeuSerGlyGlnProProPro 105
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
878  AGTTGCTTCCCTCAGTA.....ATGGGTGGGGAGGCTGCTCCAGGAG 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106  GlyThrSerGlyMetAlaProHisSerMetAlaValSerThrAlaTh 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922  GGTACAGAGGTGTGAGCCTTGCCTGAGCCCAACAGAGTACGTGTAG 971
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122  rProGlnThrGlnLeuGlnLeuGlnValAlaLeuGlnGlnGlnGln 139
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
972  CCAGGAGGACGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1021
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
139  GlnGlnGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 155
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1022  CAGATGGAGGTGGCCAGCAGATGTGGGCAATGACCTGACAGGACGCC 1071
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
156  GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGlnGlnSerAlaMetGln 172
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1072  ACACCTGGGCGCCAGCCAGACCTTATTAGCTCAGCCACTTACACAGAT 1121
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
172  nGlnPheGlnAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122  CCAGGCCCATTCATCTATTCAGACACACACACATCCACTCCAGCAGA 1171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189  GlnGlnGln.....GlnHisLeuIleLeuHisLeuHisGlnGlnGln 203
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1172  AACAGGTGTATCTCAGCAGATGTGCCATCCACAC.....CAGCAG 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204  GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1216  CAGTTCACACACCGGAGTCCAGTCCCTTCACACACGCTACACACCTCA 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219  nLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1266  GTTGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236  In.....AlaLeuGlnAlaGlnProPheGlnGlnGlnGlnProPro 248
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1316  AGCCCGAAGCCACACACCTCTACTGCTCCCTCAGCCACACAGGTCCAC 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249  MetGlnGlnProGlnProProProSerGlnAlaLeuProGlnGlnGln 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1366  ACTCAG.....CAGCTCCACACCTCCACAGTCC.....CA 1394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265  nGlnMetHisHisThrGlnHisHisGlnProProProGlnProGlnGln 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1395  GCAGCAAGCCCAACCTGTGTCTTCA..... 1422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282  rProValAlaGlnAlaGlnProSerGlnLeuProProGlnSerGlnThr 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1423  ..CCATGCTCAAGTCTTCACTTGTCTTCACTGATGACGCCCCCT 1470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299  GlnProLeuValSerGlnAlaGlnAlaLeuProGlnGlnMetLeuThr 315
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1471  AAGCCCAACCAATCC.....ATCCAAATC 1493
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
315  rGlnProProLeuLysPheValArgAlaProMetVal..... 327
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1494  CAACACCACTGTAGCTCCTATCAAGCCGCTCAGTTAGGGCCCTAGA 1543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328  .....ValGlnGlnProProValGlnProGlnValGlnGlnGln 341
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1544  TGTACAGTGGCCAGCCTACACCA.....CCCATATCCCTGTCAAGTT 1587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342  ThrAlaValGlnThrAlaGlnAlaGlnMetValAlaProGlyValGln 358
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1588  GTAGGCACTGACAGCAGGTACAGCCAGGACAGGCTTGGGGTGGC 1637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358  nValSerGlnSerSerLeuProMetLeuSerSerProSerProGlnGln 375
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1638  ACAGTGGCAGCTGTGTACT..... 1659
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
375  InValGlnThrProGlnSerMetProProProGlnProSerProGln 391
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

1660  ....ACTTCCCGGGGGATGCCAGTACAGTCAAG...TCTGGTCTAG 1698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392  ProGlyGlnProSerSerGlnProAsnSerAsnValSerGlyProAl 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1699  GCCCATTTGGCCCTTCCTGCGCCACCTTCATCCACAGCTCTGCTGCA 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408  aProSerProSerSerPheLeuProSerProSerProGlnProSerGln 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1749  GGAATGCCCTCCACATTTGCCCTGGATGATACCTTGTCTGTGTACAG 1798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425  .....SerProValThr 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1799  GGACAGCAGATGTGTAAAGGTGGGGCTTACACCTCTCCTCAGCTTGA 1848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429  AlaArgThrProGlnAsnPheSerValProSerProGlyProLeuAsn 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1849  GCCCAGTCCCTGCTGCTCTTCTATGACAGTCT.....GTGCACTT 1889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445  rProValAsnProSerSerValMetSerProAlaGlySerSerGlnAla 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1890  GCCGGGTAAACCCAGACATTTG.....GCTGTCAAAAGCAAGGCTG 1930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
462  LucGlnGlnGlnThrLeuAspLysLeuGlnLeuSerLysThrIleGlu 478
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1931  ACTGTGAGGAGGAGAGATGATGTCTCCACATTTGGGTTCAATGCTCT 1980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479  ProLeuArgArgMetIleAsnLysLysLysAsnGluAspArgLysLys 495
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1981  GCCAAAGCATCTCCAGTACGAGAAACCCAAAGTCAATGACGAGAGAG 2030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495  sAspLeu.....SerLysMetLysSerLeuLeuAspLys..... 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2031  CAGTCTTGGAGAAAGCTGAATCACTGCTAATGATGATGCTAATATCTC 2080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507  .....LeuThrAspProSerLysArgLysProLeu 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2081  CAAGCAGTGAATAGTACCTTGACC...CCGCCCCCTCAGTACGGCT 2127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517  LysThrLeuGlnLysLysGlnLeuLeuLeuGlnLysLysAsnAspM 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2128  CCACTACCTACCCATGTGT.....CTAGCAAAATGGGTGACTCAAA 2168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533  eArgGlyProLeuProHisArgPro 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2169  AC.....CCCAACAGGCCA 2182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NAL198.DAT:AAV22682
seq_documentation_block:
ID  AAV22682 standard; DNA; 2214 BP.
XX  AAV22682;
XX  28-JUL-1998 (first entry)
XX  DE  New DNA sequence isolated from Pinctada fucata.
XX  KW  Pinctada fucata; protein; cosmetic; ds.
XX  OS  Pinctada fucata.
XX  PN  JPI0080285-A.
XX  PD  31-MAR-1998.
XX  PF  28-MAY-1997; 97JP-0138461.
XX  PR  15-JUL-1996; 96JP-0184459.
XX  PA  (MIKI-) MIKIMOTO SETYAKU KK.
XX  WP1; 1998-254410/23.
XX  DR

```


XX New cDNA and e.g. vector, host cell and polypeptide - used to
 PT produce polypeptide in high yields, which is used in cosmetics
 XX
 PS Claim 1; Page 7; 15pp: Japanese.

CC The present sequence represents a new DNA sequence isolated from *Pinus*
 CC *fucata*. The polypeptide be used as an ingredient in cosmetics.
 XX

SQ Sequence 2214 BP; 393 A; 451 C; 891 G; 479 T; 0 other;

alignment_scores:

Quality: 381.50 Length: 597
 Ratio: 1.408 Gaps: 25
 Percent Similarity: 45.394 Percent Identity: 25.628

alignment_block:

US-09-668-119-3 x AAV22682/rev ..

Align seg 1/1 to reverse of: AAV22682 from: 1 to: 2214

```

50 AlaservAlserAPro.....MetasAlaleuGl 60
   |||.....|||
1631 GCGGACAGAGGATCCACCTCCACGATTTGTCTTCAAGCCCTCT 1582
   |||
60 nserLeuThGlyPro.....AlaAlaGlyAla 71
   |||
1581 AAGTGTCTTCTCCACCTGATGACAGAGGCGGCTGCACAG 1532
   |||
71 IaglyIleGlyMetProProArglyProGlyInserLeuGlyMet 87
   |||
1531 CTGGGACAGATCCCAACCAAGACCCCT..... 1503
   |||
88 GlyserPheGlyAlaMetGlyInProMetSerLeuSerGlyInPro 104
   |||
1502 .....AGTCTCCAAAGTCCACCAAAACCTCCTCC 1474
   |||
104 oProGlyThSerGlyMetAlaProHisSerMetAlaValSerThra 121
   |||
1473 ACCT.....GCACACAGCAGCAGCTAGTGCAGCGGCTAAAG 1439
   |||
121 IatHPro.GlnThrGlnLeuGlnInGlnAlaAlaLeuGlnInGln 137
   |||
1438 CTCCGCTCCGCTCCACCTCCAAAGTCCACCTCCGCTCCGACAGCG 1389
   |||
137 nGlnGlnGlnInGlnPheGlnGlnIn..... 147
   |||
1388 AGAGCAGAGCTAGAGCTCCAGCTCCACCGCGCTCCTCCTCCGTTACC 1339
   |||
148 .....GlnAlaAlaLeuGlnInGln 154
   |||
1338 ACCTCCGTTACACACCGCTCCACCTCCAGCTGAAGCTGATGCAGATG 1289
   |||
155 GlnGlnGlnGlnInGlnInGlnPheGlnAlaGln..... 166
   |||
1288 CAGATGCTGAAGCAGATGCTGAGCTTTCAATAGACCTGCAGATCCTTT 1239
   |||
167 GlnSerAlaMetGlnGlnInGlnPheGlnAlaVal..... 178
   |||
1238 AAGACATCCTCAATATGCTCAATTGAGATTTTGTGCTTGTACTGCTGA 1189
   |||
178 ..... 178
   |||
1188 TGCTTTACAGAGGCTTTTGACAGAGCTCCATTCCTCGAGAAAGTCTT 1139
   |||
178 ..... 178
   |||
1138 TAATGTCTCCATATGCTCCCTTATACACATTTGCTGAGAACGTGAA 1089
   |||
179 GlnGlnGlnGlnInGlnLeuGlnGlnInGlnInGlnIn..... 192
   |||
1088 GCAGATGACAGAGCGGCTGACAGCGGACGAGACCAAGCGGCTCCGCC 1039

```

```

193 .....HisLeuIleLysLeuHisGlnInGln 203
   |||
1038 TCTCTCTCCACCACTCCACCTCTCCGACACACAGCGGTGGGGG 989
   |||
204 GlnIleGlnGlnInGlnInGln.....LeuGl 213
   |||
988 CAGCTCAGCAGCGGACAGCAGCAGCTGCGGCTCTCTCTCTCA 939
   |||
213 nArgIleAlaGlnLeuGlnInGlnInGln..... 224
   |||
938 CCGGCTCTACCTCTCTCCACCTGTAAGAAACCAACACAGCGATCTCC 889
   |||
225 .....GlnGlnGlnGlnInGlnInGlnInGlnInGlnAlaLeu 238
   |||
888 ACCTCGCGACAGACCGGCGGACAGCAGCAGCGGACGAGATCTCTC 839
   |||
838 CATAT.....CTCCGAGGCGCACCAAGACCTCCGAGGCTCCAAATCCGCC 793
   |||
255 oProSerGlnAlaLeuProGlnInGlnInGlnInGlnHisThrGln 272
   |||
792 ACCGAGTCTCTCA..... 780
   |||
272 IHisGlnProProGlnProGlnInGlnProProValAlaGlnInGln 288
   |||
779 .....AGTCCGCCACCGAGTCTCCAAATGCACTGCG..... 747
   |||
289 ProSerGlnLeuProProGlnInGlnInGlnProLeuValSerGlnAl 305
   |||
746 .....CTCTCTGACAGGGGACCA.....GCGGCTGGCGC 718
   |||
305 aGlnAlaLeuProGlyGlyMetLeuThrGlnProProLeuLysPhe 322
   |||
717 TGCAGCAGAGCTCTCTCAAGTCTACCTCGGCTCTCTGACAGCGGACG 668
   |||
322 aAlaArgAlaProMetValAlaGlnInGlnProProValGlnInGln 338
   |||
667 CACCGGCTGGGCTGGGCGACAGCTCCACCACTCCACCCCGGCG... 621
   |||
339 GlnGlnGlnThrAlaValGlnThrAlaGlnAlaAlaGlnMetValAla 355
   |||
620 .....GACAGACGGGAGCTGCTCCAGCAGCAGCAGCAGCTCCAGCTCC 577
   |||
355 oGlyValGlnValSerGlnSerLeuProMetLeuSerProSerP 372
   |||
576 ACCAGCTCCACCGAGGCGCAAGAGCTGCACCAAGA...TCATGAAATCTA 530
   |||
372 roGlyGlnInGlnValGlnThrProGlnSerMetPro...ProProGln 387
   |||
529 AATCAATTAATATGCTCTCCACCGCAAGTCCGCGCAAGACTCCAACTCG 480
   |||
388 ProSerProGlnProGlyInProSerSerGlnProAsnSerAsnValSe 404
   |||
479 CCAAGTCCACCTCCGAGACCTCG..... 456
   |||
404 rSerGlyProAlaProSerProSerSerPheLeuProSerProSerPro 421
   |||
455 .....CCCAATCCAAAGTCCGAGT.....CCAGACGCTGCTCCG 422
   |||
421 InProSerGlnInserProValThrAlaArgThrProGlnAsnPheSerVal 437
   |||
421 CTCTGCGGCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 372
   |||
438 ProSerProGlyProLeuAsnThrProValAsnProSerSerValMetSe 454
   |||
371 CAGACACACCTCCGCGGCGGCTCCATTCACCTGATCCATCCAGTCC 322
   |||
454 rProAlaGlySerSer.GlnAlaGlnGlnInGlnInGlnLysPheLeu 470
   |||
321 CCATTCATCATGCTCATGCTGTCGTCATCATCATGATCATTCATTCATTC 272

```

```

471 |LysGlnLeuSerLysYrIleGluProLeuAtrgMetIleAsnLysII 487
    |:::|
271 |CACCC.....CGCCGCCGCTCCG..... 251
    |
487 |AspLysAsnGlnAspArgLysLysAspLeuSerLysMetLysSerLeuL 504
    |:::|
250 |.....CGCGGCTCC 241
    |
504 |euAspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
    |:::|
240 |PCCAGACCCGCGAGCCGCCAGAGCGCTCCGCCCTCCGACAC..... 197
    |
521 |LysCysGluIleAlaLeuGluLysLeuLysAsnAspMetArgCysProLe 537
    |:::|
196 |.....CGTAAAGACCTGCACCTGCTCAT 174
    |
537 |uPro.....HisArgProArgCysHisArgPro 546
    |:::|
173 |GCCGAACGCAACCAATGCCCCATGATCAATAACCA 137
    |
seq_name: /SID22/9/cgdata/geneseq/9/geneseq/NA1998.DAT:AAV22683
seq_documentation_block:
ID AAV22683 standard; cDNA to mRNA; 3331 BP.
XX
AC AAV22683;
XX
DT 28-JUL-1998 (first entry)
XX
DE New DNA sequence isolated from Pinctada fucata.
XX
KW Pinctada fucata; protein; cosmetic; ds.
XX
OS Pinctada fucata.
XX
FH Key 1 Location/Qualifiers
FT CDS 50..2266
   /*leg= a
XX
PN JPI0080285-A.
XX
PD 31-MAR-1998.
XX
PF 28-MAY-1997; 97JP-0138461.
XX
PR 15-JUL-1996; 96JP-0184459.
XX
PA (MIKI-) MIKIMOTO SEIYAKU KK.
XX
DR WPI; 1998-254410/23.
XX
DR P-PsDB; AAW56163.
XX
PT New cDNA and e.g. vector, host cell and polypeptide - used to
PT produce polypeptide in high yields, which is used in cosmetics
XX
PS Claim 2; Pages 7-9; 15pp; Japanese.
XX
CC The present sequence represents a new DNA sequence isolated from Pinctada
CC fucata. The encoded polypeptide be used as an ingredient in cosmetics.
XX
SQ Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

```

alignment scores:

Quality: 381.50 Length: 597
 Ratio: 1.408 Gaps: 25
 Percent Similarity: 45.394 Percent Identity: 25.628

alignment block:

us-09-668-119-3 x AAV22683/rev ..
 Align seg 1/1 to reverse of: AAV22683 from: 1 to: 3331

```

50 |AlaSerValSerAspPro.....MetAsnAlaLeuGI 60
    |:::|
1680 |GCGGACAGGAGATCCACCTCCACGCAATTTGTTCTGTAAGAGCTCTT 1631
    |
60 |nSerLeuThrGlyIlePro.....AlaIleGlyAla 71
    |:::|
1630 |AACTGCTCTTCTCCACACCTGATGACAGCGGCGGCTGACAGCAG 1581
    |
71 |IleGlyIleGlyMetProProArgGlyProGlyGlnSerLeuGlyIleMet 87
    |:::|
1580 |CTCGGCGAGATCCGCCACCAAGACCGCT..... 1552
    |
88 |GlySerPheGlyAlaMetGlyGlnPrometSerLeuSerGlyIleProPr 104
    |:::|
1551 |.....AGTCCTCCAGTCACCAAAACCTCTCC 1523
    |
104 |oProGlyThrSerGlyMetAlaProHisSerMetAlaValAlaSerThr 121
    |:::|
1522 |ACCT.....GCACCAAGCAGCAGCTAGTGCAGGCGGTAAG 1488
    |
121 |LeuThrProGlnThrGlnLeuGlnLeuGlnGlnAlaAlaLeuGlnGln 137
    |:::|
1487 |CTCGGCTCTCGGCTCCACCTCCAGTCACCTCCGCTCTGACAGCGC 1438
    |
137 |ngGlnGlnGlnGlnGlnPheGlnGlnGln..... 147
    |:::|
1437 |AGAGCAGCAGCTAGAGCTCCAGCTCCACGCGCGCTCTCTCCCTTACC 1388
    |
148 |.....GlnAlaAlaLeuGlnGln 154
    |:::|
1387 |ACCTCGTTACCAACACCGCTCCACCTCCAGCTCAAGAGCTGATGAGATG 1338
    |
155 |GlnGlnGlnGlnGlnGlnGlnGlnPheGlnAlaGln..... 166
    |:::|
1337 |CAGATGCTGAGCAGATGCTGAGCTTTTCATGATGACCTGCAAGATCTTT 1288
    |
167 |GlnSerAlaMetGlnGlnGlnPheGlnAlaValAla..... 178
    |:::|
1287 |AAGACATCTCTCAAAATGCTCAATTTGAGATTTTGCTGCTGACTAGTGA 1238
    |
178 |..... 178
    |
1237 |TGCTTTAGCAGAGGCTTTTCAGAGGCTCCATTACTCTCGAAGAGCTTT 1188
    |
178 |..... 178
    |
1187 |TAATGCTCTTAATGCGCTCCCTTATACCACTGATTTGCTAGAGCTGA 1138
    |
179 |GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGln..... 192
    |:::|
1137 |GCAATGACAGCAGCGGCTCCAGCGGCGGCGGCGGCTCCGCC 1088
    |
193 |.....HisLeuIleLysLeuHisGlnGlnGln 203
    |:::|
1087 |TCCTCTCCACCACTCCACCTCTCCGCGCACACGAGCGGCTGCGGCG 1038
    |
204 |GlnIleGlnGlnGlnGlnGln.....LeuGI 213
    |:::|
1037 |CAGCTGACAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCTG 988
    |
213 |nArgIleAlaGlnLeuGlnLeuGlnGlnGlnGln..... 224
    |:::|
987 |CGGCTCTACCTCTCTGTCACACCTAGAAACCAACACACGAGCTCTCC 938
    |
225 |.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeu 238
    |:::|
937 |ACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCTG 888
    |
239 |GluAlaGlnProIleGlnGlnProMetGlnGlnProGlnProPr 255
    |:::|
887 |CATAT.....CTTCGAGGCGCACCAAGACCTCCGAGCGCTCCAAAGTCGCG 842
    |
255 |oProSerGlnAlaLeuProGlnGlnLeuGlnGlnGlnGlnGlnGln 272

```

```

841 ACCGAGTCTCCA..... 829
272 ishtsglnProProglInProglInProValAlaInaInglIn 288
828 ....AGTCGGCCACCGAGTCTCCCAAGTCCACCTCCG..... 796
289 ProSerGlnLeuProProglInProglInProLeuValSerGlnAl 305
795 .....CTCTCGACGGCGACGA.....GCGCTCCGCG 767
305 aglnAlaLeuProglInMetLeuYrThrGlnProProLeuLysPheY 322
766 TGCAGCAGAGCTCTCCCAAGTCTCCCTCTCTCGACGGCGCAG 717
322 alArgAlaProMetValValGlnInProProValAlaInProglInValGln 338
716 CAGCGGCTGGCGCTGGCGGACAGCTCCACCACTCCACCCCGGCG... 670
339 GlnGlnInThrAlaValGlnThrAlaGlnAlaGlnMetValAlaPr 355
669 .....CAGCAGCGGCGAGCTGACAGCAGCAGCAGCAGCTCCAGCTCC 626
355 oglyValGlnValSerGlnSerLeuProMetLeuSerSerProSerP 372
625 ACCAGCTCCACCGAGGCAAGAGCTGCACCAAGA...TCATCGAATCTA 579
372 roglyGlnInValGlnThrProGlnSerMetPro...ProProProGln 387
578 AATCAATTAATATGCTCTCCACCGCAGTCCGCCAAGACCTCCCAAGTCG 529
388 ProSerProGlnProGlnInProSerSerGlnProAsnSerValSe 404
528 CCAAGTCCACCTCCGAGACCTCCG..... 505
404 rserGlyProAlaProSerProSerSerPheLeuProSerProSerProg 421
504 .....CCCAATCCAAAGTCCGAGT.....CCAGCAGCTCTCTCTG 471
421 InProSerGlnSerProValThrAlaArgThrProGlnAsnPheSerVal 437
470 CTCTGGGCGCTCTCCAGCAGCAGCAGCAGCAGCTCCGCGACCA 421
438 ProSerProGlnProLeuAsnThrProValAsnProSerSerValMetSe 454
420 CCAGCAGCAGCTCCGCGGCGCAGCTCCATCCAGTATCCTCCAGTCTC 371
454 rProAlaGlySerSerGlnAlaGlnGlnGlnInThrLeuAspLysLeu 470
370 CCATTCAATCATGTCATCGCTGCTCATCATATATCCCAATCTC 321
471 LysGlnLeuSerLysTyrIleGlnProLeuArgArgMetIleAsnLysII 487
320 CACGCG.....CGCGCGCGCGTCCCGC..... 300
487 easpLysAsnGlnLysPargLysLysAspLeuSerLysMetLysSerLeuL 504
299 .....GTCGCTCC 290
504 euAspIleLeuThrAspProSerLysArgGlyProLeuLysThrLeuGln 520
289 TCCACGACACCGGCGAGCTCCCGAGCGCGCTCCGCCCTCCGACAC..... 246
521 LysCysGluIleAlaLeuGlnLysLeuLysAsnAspMetArgCysProLe 537
245 .....CGTAAGACCTGCTCAT 223
537 uPro.....HisArgProArgCysHisArgPro 546
222 GCCGAAGACGACACCATGCCATGATCATTAACCA 186
seq_name:/SIDS2/gcdata/geneseq/geneseq/NA1999.DAT.AA232021

```

```

seq_documentation_block:
ID AA232021 standard; DNA: 3331 BP.
XX AC AA232021;
XX AC 10-JAN-2000 (first entry)
XX DE Human METH1 related EST D86074.
XX KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
XX KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
XX KW angiogenesis inhibitor; abnormal wound healing; inflammation;
XX KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
XX KW diabetic retinopathy; macula degeneration; haemangioma; detection;
XX KW arterial-venous malformation; immune deficiency; ss.
XX OS Homo sapiens.
XX PN W09937660-A1.
XX PD 29-JUL-1999.
XX PF 22-JAN-1999; 99WO-US01313.
XX PR 23-JAN-1998; 98US-0072298.
XX PR 28-AUG-1998; 98US-0098539.
XX PA (IRUE/) IRUELA-ARISPE L.
XX PA (HAST/) HASTINGS G A.
XX PI (RUBE/) RUBEN S M.
XX PI IrueLa-Arispe L, Hastings GA, Ruben SM;
XX DR WPI: 1999-590684/50.
XX PT New isolated metalloprotease thrombospondin polypeptides, useful for
XX PT treating hyperproliferative disorders, cancers or autoimmune disorders
XX PT -
XX PS Disclosure: Page 322-324; 457pp; English.
XX CC AA232000 and AA232001 encode, and AAY49501 and AAY49502 represent, human
XX CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
XX CC respectively. METH1 and METH2 have been found to be potent inhibitors of
XX CC angiogenesis both in vitro and in vivo. They can be used for treating
XX CC cancer and other disorders related to angiogenesis including abnormal
XX CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX CC endometrial bleeding disorders, diabetic retinopathy, some forms of
XX CC macula degeneration, haemangiomas, and arterial-venous malformations.
XX CC They may be useful in treating deficiencies or disorders of the immune
XX CC system, by activating or inhibiting the proliferation, differentiation,
XX CC or mobilisation (chemotaxis) of immune cells. The etiology of these
XX CC immune deficiencies or disorders may be genetic, somatic, such as
XX CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
XX CC toxins), or infectious. They can also be used to treat inflammatory
XX CC conditions, both chronic and acute conditions. The products can also be
XX CC used for detection and diagnosis. AA232002 to AA232080, and AAY49503 to
XX CC AAY49511 represent sequences given in the exemplification of the present
XX CC invention.
XX SQ Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

alignment_scores:
Quality: 381.50 Length: 597
Ratio: 1.408 Gaps: 25
Percent Similarity: 45.394 Percent Identity: 25.628

alignment_block:
US-09-668-119-3 x AA232021/rev ..
Align seg 1/1 to reverse of: AA232021 from: 1 to: 3331

```

```

50 AlaSerValSerAspPro.....MetAsnAlaLeuGI 60
1680 GCGGAGAGAGGGGATACCTCCACAGCATTTGCTCTCAAGGCCCTCT 1631
6 nserLeuThrGlyG/Pro.....AlaAlaGlyAlaA 71
1630 AAGTCTCTTCTCTCCACCACCTGATGACGACGGGACGGCTCGACGAG 1581
71 laGlyIleGlyMetProProArgGlyProGlyGlnSerLeuGlyGlyMet 87
1580 CTGCGGACGATCCCTACCAAGACCGCCT..... 1552
88 GlySerPheGlyAlaIleGlyGlnPrometSerLeuSerGlyGlnProPr 104
1551 .....AGTCTCTCAAGTCCACCAAAACCTCTCC 1523
104 oProGlyThrSerG/MetAlaProHisSerMetAlaValaSerThrA 121
1522 ACCT.....GCACACGACGACGTAAGTGCACGCGCTAAAG 1488
121 laThrPro.GlnThrGlnLeuGlnLeuGlnValaAlaLeuGlnGlnI 137
1487 CTGCGGCTCCGCGCTCCACCTCAAGTCCACCTCGGCTCTCGACGACG 1438
137 nGlnGlnGlnGlnPheGlnGlnGln..... 147
1437 AGAGCAGACGAGTAGACCTCCAGCTCCACGCGGCTCTCTCGTACC 1388
148 .....GlnAlaAlaLeuGlnGlnGln 154
1387 ACCTCCGTTACACCA/CCGCTCCACCTCCAGCTGAAGCTGATGCAGATG 1338
155 GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGln..... 166
1337 CAGATGCTGAAGCAGTGTGACGCTTTCAATAGACCTGCAGATCTTT 1288
167 .GlnSerAlaMetG/GlnGlnPheGlnAlaVala..... 178
1287 AAGACATCTTCAATGCTCAATTTGAGATTTGCTTCTTACTGTGA 1238
178 ..... 178
1237 TGCTTAGCAGAGGCTTTTGACAGGCTCATTACTCTAGAAAGCTT 1188
178 ..... 178
1187 TAATGCTCTTAATGCTGCCCTTATACCACTCATTTGCTAGAAAGCTGA 1138
179 .GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGln..... 192
1137 GCAGATGCACACGCGCTGCACGCGGACGCGGACACACGCGCTCGCC 1088
193 .....HisLeuIleLeuHisHisGlnAsnGlnGln 203
1087 TCTCTCCACACCTCCACCTCTCTCGCCACACGCGGCTCGCGCG 1038
204 GlnIleGlnGlnGlnGlnGln.....LeuGI 213
1037 CAGCTGAGCAGCGGACGACGACGACGCTGTGGGCTCTCTCTCTCA 988
213 naGlyIleAlaGlnLeuGlnLeuGlnGln..... 224
987 CCGCGTCTACCTCTCTCTCGCGGAAACCAACCAACGAGTCTCTCC 938
225 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 238
937 ACCTCGGCGACGACGCGCGGACGACGACGACGACGACGACGATCTCTC 888
239 GlnAlaGlnProProIleGlnGlnProMetGlnGlnProGlnProPr 255
887 CATAT.....CTTCCGAAGCCACCAAGCTCTCGAGGCTCTCAAGTCCGCC 842
255 loProSerGlnAlaLeuProGlnGlnLeuGlnMetHisHisThrGlnH 272
841 ACCGAGTCTCCA..... 829
272 tshisGlnProProProGlnProGlnGlnProValaAlaGlnAsnGln 288
828 .....AGTCCGACCGAGTCTCTCAAGTCCACCTCG..... 796
289 ProSerGlnLeuProProGlnGlnGlnProValaSerGlnAl 305
795 .....CTCTCTCAGCGGACGA.....GGGCTCGGCG 767
305 aGlnAlaLeuProGlyGlnMetLeuThrGlnProProLeuLysPheV 322
766 TGACGACGAGCTCTCCCAAGTCTACCTGCGCTCTCTCGACGCGCAG 717
322 aAlaGlnAlaPrometValaGlnGlnGlnProValaGlnProGlnValaGln 338
716 CAGCGGCTCGGCTCGGCGGACGACGCTCCACCACTCCACCCCGCG... 670
339 GlnGlnGlnThrAlaValGlnThrAlaGlnAlaGlnMetValaLap 355
669 .....GCAGCAGCGGACGCTGACGACGACGACGACGACGACGCTCC 626
355 oGlyValGlnValSerGlnSerSerLeuPrometLeuSerProSerP 372
625 ACCAGCTCCACGCGGACGAGGACGACGACGACGACGACGACGACGACG 579
372 roGlyGlnGlnValGlnThrProGlnSerMetPro...ProProGln 387
578 AATCAATTAATGCTCTCCACCGCCACGACGACGACGACGACGACGACG 529
388 ProSerProGlnProGlyGlnProSerSerGlnProAsnSerAsnValSe 404
528 CCAAGTCCACCTCCGACCTCG..... 505
404 rserGlyProAlaProSerProSerSerPheLeuProSerProSerProg 421
504 .....CCCAATCCAGTCCGAGT.....CCACACCTGCTCTGTG 471
421 InProSerGlnSerProValaThrAlaArgThrProGlnAsnPheSerVal 437
470 CTCTGCGGCTCTCTCCAGCACGACGACGACGACGACGACGACGACGAC 421
438 ProSerProGlyProLeuAsnThrProValaAsnProSerSerValMetSe 454
420 CCAGCACCACTCCGCGGCGGACGCTCCATTCACCGTATCATCCAGTC 371
454 rProAlaGlySerSer.GlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 470
370 CATTCATCATGCTGATGCTGCTCATCATCATCATCATCATCATCATCATC 321
471 LysGlnLeuSerLysTyrIleGlnProLeuArgArgMetIleAsnLysII 487
320 CACCGC.....CGCGCGCGCTCGCGC..... 300
487 easPlyAsnGlnLysPargLysAspLeuSerLysMetLysSerLeuL 504
299 .....CGTGGGCTC..... 290
504 euAspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
289 TCCAGCACCGGACGCTCCCGGCGGCTCGCGGCTCTCGACAC..... 246
521 LysCysGlnIleAlaLeuGlnLysLeuLysAsnAsnMetArgCysProLe 537
245 .....CGTAAACCTCGACGCTCTCAT 223
537 uPro.....HisArgProArgCysHisArgPro 546
222 GCCGAACCGCAACCAATGCCCATGATCAATAACCA 186

```

Fri, Mar 1 09:12:02 2002

us-09-668-119-3.rng

Page 17




```

922 GSTACAGAGTGTGTGAGCCCTTGGCTGACGCCAAGCATGTGTGTGAG 971
122 rProglInthrGlnLeuGlnGlnValAlaLeuGlnGlnGlnGlnGln 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
972 CCAGGGGAGCCAGACAGAGGCAAGAGTGCAGACGCAAGAAAGGACGAG 1021
139 InGlnGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 155
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1022 CAGATGGAGGTGGCCAGCAGATGGGCAATGAGCCTGACAGACGACGCC 1071
156 GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGlnGlnGlnGlnGlnGln 172
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1072 ACACCTGGCCGAGCAGACCTTATGTAGCTCAGCCACTTACACAGCAT 1121
172 nGlnPheGlnAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 189
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1122 CCAGGCCCATTCACATTCATTGAGCAACAGCAACAGATCCACTCAGCAG 1171
189 InGlnGln..... GlnHisLeuIleuLysLeuHisGlnGlnGlnGln 203
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1172 AACAGGTGTGTATCCAGCAGAGATGGCATCCACCC.....CAGCAG 1215
204 GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 219
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1216 CAGTTCACGACCGGAGTCCAGCTCCTTCCACACAGCTTACACACCTTCA 1265
219 nLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 236
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1266 GTTGGCGCAGCAGACACAGCAGCAACACAGCAGCAGCAGCAGCAGCAG 1315
236 In.....A.....AlaLeuGlnAlaGlnProPheIleGlnGlnProPro 248
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1316 AGCCGCAAGCCACCAACCTCTACTGCTCCCTCAGCCACACAGAGTCCACCT 1365
249 MetGlnProGlnProPheProSerGlnAlaLeuProGlnGlnGlnGln 265
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1366 ACTCAG.....CAGTCCACCTTCCACAGTCC.....CA 1394
265 nGlnMetHisIleuGlnHisIleuGlnProProProGlnProGlnGln 282
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1395 GCAGCAAGCCCAACCCGTGCTGTGAG..... 1422
282 rOProValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 298
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1423 ..CCATGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1470
299 GlnProLeuValSerGlnAlaGlnAlaLeuProGlnGlnGlnGlnGln 315
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1471 AAGCCACCAATCC.....ATCCATC 1493
315 rGlnProProLeuLysPheValAlaGlnAlaProMetVal..... 327
1494 CAACCAACCTGTAGC/CCTTTCAGCGCGCTGATTGAGGCGCGCTAAGA 1543
328 .....ValGlnGlnProProValGlnProGlnGlnGlnGlnGlnGln 341
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1544 TGTACGCTGCCAGCAGCAACCA.....CCCATATCCTGTGCAAGTT 1587
342 ThrAlaValGlnThrIleGlnAlaGlnMetValAlaProGlnValGln 358
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1588 GTAGGCACTGACAGCAGGTACAGCCAGCAGCAGCTTGTGGGTGGC 1637
358 nValSerGlnSerSerLeuProMetLeuSerSerProSerProGlnGln 375
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1638 ACAGCTGGCAGCTGTGTACT..... 1659
375 InValGlnThrProGlnSerMetProProProProGlnProSerProGln 391
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1660 .....ACTTCCCGGGGATCCAGGTACAGTCAAG...TGTGCTCAG 1698
392 ProGlnGlnProSerGlnProSerGlnProSerGlnValSerSerGlnPro 408
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1699 GCCCATTTGGCTCTGCGCCACCTTATCCAGGCTCTGTGCTGCACTGCA 1748

```

```

408 aProSerProSerSerPheLeuProSerProSerProGlnProSerGln 424
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1749 GAGTGGCTCCACCATGTGGCCCTGGATGACCTTGTCTGCTGTGACG 1798
425 .....SerProValThr 428
1799 GAGACAGCATGTGTAAAGGTGGGCTACACCTCTCCTGCTGTGTA 1848
429 AlaArgThrProGlnSerPheSerValProSerProGlnProLeuAsn 445
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1849 GCCCAGGTCTCTGCTGCTTATATGCACT.....GTGACTT 1889
445 rProValAsnProSerSerValMetSerProAlaGlnSerSerGlnAlaG 462
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1890 GCGGGTAAACCCAGACATG.....GCTGTCAAGCCAGGCTG 1930
462 InGlnGlnGlnIleuAspLysLeuLysGlnLeuSerLysTyrIleGln 478
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1931 ACTGTAGGAGGAGAGATGATGTCTCCACATTTGGCTTCAATGCTCT 1980
479 ProLeuAlaArgMetIleAsnLysIleAspLysAsnGlnAspArgLys 495
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1981 GCCAAAGCATCTCCAGTACAGACAGAACCCAAAGATCATGACGAGAGAG 2030
495 AspLeu.....SerLysMetLysSerLeuLeuAspIle..... 506
2031 CAGTCTTGGAAGAAAGCTGATCATGCTGCTATGCTATGCTATGCTATGCT 2080
507 .....LeuThrAspProSerLysArgLysProLeu 516
2081 CAAGCAGTGAATGATGAGCTTGACC...CCGCCCCCTTCACTACGCT 2127
517 LysThrLeuGlnLysCysGlnIleAlaLeuGlnLysLeuLysAsnAsp 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2128 CCACTGCTGAGCTGT.....CTAGACAAAGGGGTGACTCAAA 2168
533 eLargCysProLeuProHisArgPro 541
2169 AC.....CCCCAGAGGCCA 2182

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-08-864-038A-1
seq_documentation_block:
Sequence 1, Application US/08864038A
GENERAL INFORMATION:
APPLICANT: KUNIO NAKASHIMA et al.
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: 812-5 HIRANO
STREET: IISHINDEN
CITY: TSU-CITY
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette, 3.50 inch, 1.44 MB storage
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:

```


NAME: C. Bruce Hamburg
 REGISTRATION NUMBER: 22,389
 REFERENCE/DOCKET NUMBER: F-5610
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)986-2340
 TELEFAX: (212)953-7733
 INFORMATION FOR SEO ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2214
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 ORIGINAL SOURCE:
 ORGANISM: Pinclada fucata
 CELL TYPE: mantle epithelial cell
 US-08-864-038A-1

Alignment scores:
 Quality: 381.50 Length: 597
 Ratio: 1.408 Gaps: 25
 Percent Similarity: 45.394 Percent Identity: 25.628

Alignment block:

US-09-668-119-3 x US-08-864-038A-1/rev ..

Align seg 1/1 to reverse of: US-08-864-038A-1 from: 1 to: 2214

50 AlaservAlaservPro.....Metasnlaleu1 60
 |||.....
 1631 GCGGACAGAGGATCCACCTCCACGATTTGTCTTCAAGCCCTCT 1582
 60 nserleutherglyglyPro.....AlaAlaGlyAla 71
 |||.....
 1581 AAGTCTCTCTCTCCACCTGATGCAGCAGCGCGCTGCTCAGCAG 1532
 71 laglylleglymetProProarglyProglylserleuglyglyMet 87
 |||.....
 1531 CTGGCGGAGATCCCGCCACAGACCGCT..... 1503
 88 glyserpneglyAlaMetglylGlnProMetserleuSerlyGlnPro 104
 |||.....
 1502AGTCTCCAGATCCACCAAAACCTCTCC 1474
 104 oProglylthserglyMetAlaProHlserMetAlaValaSerthra 121
 |||.....
 1473 ACCT.....GCACACAGCAGCAGTGTGACACGCGCTAAAG 1439
 121 latProGlnThnGlnleuGlnleuGlnAlaAlaGlnGlnGln 137
 |||.....
 1438 CTGGCGCTCCCTCCACCTCCAAAGTCCACCTCGGCTCTGACAGCAG 1389
 137 nGlnGlnGlnGlnGlnpneGlnGlnGln..... 147
 |||.....
 1388 AGAGCAGCAGCTAGAGCTCCAGCTCCACCGCGCTCTCTCCGTTACC 1339
 148GlnAlaAlaLeuGlnGln 154
 |||.....
 1338 ACCTCCGTTACCCACCGCTCCACCTCCAGCTGCAAGCTGATGACATG 1289
 155 GlnGlnGlnGlnGlnGlnGlnGlnpneGlnAlaGln..... 166
 |||.....
 1288 CAGATGCTGAGCAGATGCTGACTTTTCAATAGACCTGCAAGATCTTT 1239
 167 GlnserAlaMetGlnGlnGlnpneGlnAlaVala..... 178
 |||.....
 1238 AAGACATCTTCAAAATCGTCAATTTGATTTGTGCTTGTCTACTGCTGA 1189
 178 178
 1188 TGCTTACGAGAGCTTTTGAGAGCTCCATTTACTCTGAGAGGCTCT 1139
 178 178

1138 TAATGCTCTTAATGCTCTCTTATACCATCATTTGTAGAACTGAA 1089
 179 GlnGlnGlnGlnGlnleuGlnGlnGlnGlnGlnGlnGln..... 192
 |||.....
 1088 GCAGTGCAGCAGCGGCTGACAGCGGCTGACAGCAGCGGCTCTCCGCC 1039
 193HisleuIleuHisleuHisleuGln 203
 |||.....
 1038 TCTCTCTCCACCATCTCCACCTCTCTCCGACACACAGCGGCTGGGCG 989
 204 GlnIleGlnGlnGlnGlnGln.....LeuG1 213
 |||.....
 988 CAGCTCAGCAGCGGACGAGCAGCAGCTGCGGCTCTCTCTCTGCA 939
 213 nArgIleAlaGlnleuGlnleuGlnGln..... 224
 |||.....
 938 CCGCTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 889
 225GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnleu 238
 |||.....
 888 ACCTCGGACAGCAGCGGCTGACAGCAGCAGCGGCTGACAGCAGATCTC 839
 239 GluAlaGlnProProIleGlnProProMetGlnGlnProGlnPro 255
 |||.....
 838 CATAT.....CTCCGAGGCTCCAGACCTCCGAGGCTCTCTCTCTCT 793
 255 oProSerGlnAlaLeuProGlnGlnleuGlnleuHisIleGln 272
 |||.....
 792 ACCGAGCTCTCA..... 780
 272 IshIsgInProProProGlnProGlnGlnProProValaGlnIn 288
 |||.....
 779AGTCCGACCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 747
 289 ProSerGlnleuProProGlnProGlnGlnProleuValaSerGln 305
 |||.....
 746CTCTCTGAGCGGACGCA.....GCGGCTGCGGC 718
 305 aglnAlaLeuProGlylGlnMetleuTyThrGlnProProLeuLysPhe 322
 |||.....
 717 TGCAGCAGCAGCTCTCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 668
 322 alArgAlaProMetValaGlnGlnProProValaGlnProGlnVala 338
 |||.....
 667 CAGGCTCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 621
 339 GlnGlnGlnThrAlaValGlnThrAlaGlnAlaAlaGlnMetVala 355
 |||.....
 620GCAGCAGCGGCTGAGCTGAGCAGCAGCAGCAGCAGCTCTCC 577
 355 oglyValGlnValserGlnserSerleuProMetleuSerProSer 372
 |||.....
 576 ACCAGCTCCACGAGGCAAGAGCTGACCAAGA...TCATCGAAATCTA 530
 372 roGlyGlnGlnValGlnThrProGlnSerMetPro...ProProGln 387
 |||.....
 529 AATTAATTAATGCTCTCCACCGGCAAGTCTGCGCAAGCTCCAGTCCG 480
 388 ProSerProGlnProGlyGlnProSerSerGlnProSerValaSe 404
 |||.....
 479 CCAAGCTCCACCTCCGAGACCTCG..... 456
 404 rserGlyProAlaProSerProSerSerPheleuProSerProSer 421
 |||.....
 455CCAATCTCAAGTCCGAGT.....CCAGCAGCTCTCTCT 422
 421 InProSerGlnserProValaThrAlaArgThrProGlnAsnPheserVal 437
 |||.....
 421 CTCTGCGGCTCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 372
 438 ProSerProGlyProleuAsnThrProValaAsnProSerSerValMet 454
 |||.....

```

371 CCAGCACCACCTCGG JCGCAGCTCCCATTCACCGTCATCATCCAGTC 322
454 rProLaglySerSer GlnAlaGlnGlnGlnGlnTyrLeuAspLysLeu 470
1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
321 CCATTGATCATGCTCTGCGCTGCTCATGATCATATTCATCCCATCTC 272
471 LysGlnLeuSerLysTyrIleGlnProLeuArgMetIleAsnLysII 487
271 CACCC.....GCGCGCGCGCTCCG..... 251
487 eAspLysAsnGlnAspArgLysLysAspLeuSerLysMetLysSerLeu 504
250 .....GCTGCGCTCC 241
504 euAspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 TCCAGCAGCCGCGAGCTCCCGCAGCGCTCCGCGCTCCGACAC..... 197
521 LysCysGlnIleAlaLeuGlnLysLeuLysAsnAspMetArgCysProLe 537
196 .....CGTAAGACCTGCACTGCTCAT 174
537 uPro.....HisArgProArgCysHisArgPro 546
173 GCCGAACGCAACCAATGCCCCAGATCATTAACCA 137
seq_name: /cgn2_6/ptodact/2/lna/bA_COMB.seq:US-08-864-038A-2

```

```

seq_documentation block:
; Sequence 2, Application: US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: KUNIO, NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinder
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3331
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Pinclada fucata

```

```

; CELL TYPE: mantle epithelial cell
; FEATURE: mRNA
; LOCATION: from 1 to 3331
; IDENTIFICATION METHOD: E (by experiment)
; US-08-864-038A-2

alignment_scores:
    Quality: 381.50      Length: 597
    Ratio: 1.408         Gaps: 25
    Percent Similarity: 45.394      Percent Identity: 25.628

alignment_block:
    US-09-668-119-3 x US-08-864-038A-2/rev ..

Align seg 1/1 to reverse of: US-08-864-038A-2 from: 1 to: 3331

50 AlaSerValSerAspPro.....MetAsnAlaLeuG 60
1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1680 GCGGCAAGCGGAGATCCACCACCTGATGACCGCATTTGTTCTTAAGCCCTTCT 1631
60 nSerLeuThrGlyPro.....AlaAlaGlyAla 71
1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1630 AAGTGTCTTCTCTCCACACCTGATGACAGCAGCGGAGCGGCTGACGACAG 1581
71 IaGlyIleGlyMetProProArgGlyProGlyGlnSerLeuGlyMet 87
1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1580 CTGCGGCAAGATCCCGCCACAGACCGCT..... 1552
88 GlySerPheGlyAlaMetGlyInPromeSerLeuSerGlyGlnPro 104
1551 .....AGTCTCCAGTCCACCAACCAACCTCTCC 1523
104 oProGlyThrSerGlyMetAlaProHisSerMetAlaValSerTrp 121
1522 ACCT.....GCACCAAGCAGCAGTGTGAGCGGCTAAG 1488
121 IaThrPro-GlnThrGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 137
1487 CTGCGGCTCGGCTCCACCTCCAGTCAATGCACTCCGCGCTCTGACACAGG 1438
137 nGlnGlnGlnGlnGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGln 147
1437 AGAGCAGCAGCTAGAGCTCAGCTCCACGCGGCTCTCTCTCTCTCTCTCTCT 1388
148 .....GlnAlaIleuGlnGlnGln 154
1387 ACCTCGTTACCAACGCGCTCCACCTCCAGCTGAGTGAATGATGAGATG 1338
155 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 166
1337 CAGATGCTGAGCAGATGCTGAGCTTTCAATAGACCTGCAAGATCTTTT 1288
167 GlnSerAlaMetGlnGlnGlnGlnPheGlnAlaVal..... 178
1287 AAGACATCTTCAATGTCATATTGAGATTGTTGCTGTCTACTGCTGA 1238
178 ..... 178
1237 TGCTTACAGAGGCTTTTGACAGAGCTCCATTACTCTGGAAGGCTT 1138
178 ..... 178
1187 TAATGTCCTTAATGCGCTTATACCATTCATTTGTCTAGAAAGCTGA 1138
179 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 192
1137 GCAATGACAGCGGCTGACGCGGCGGCGGCGGCGGCGGCTCCGCC 1088
193 .....HisLeuIleLysLeuHisGlnGlnGlnGlnGlnGlnGln 203
1087 TCTCTCTCCACACCTGACCTCTCTCGGCGGCGGCGGCGGCGGCGGCGG 1038

```

[illegible]

```

299 .....::: |||||
504 euAsp1leuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
||::: ||| :::: |||||:::
289 tccAcacacccGccGctccGccAgccGctccGccGctccGcAc. .... 246
521 LysCysGlu1leAlaLeuGlnLysLeuLysAsnAspMetArgCysProLe 537
245 .....::: |||||::: cctAAAGAcctGcAcTgcTccAT 223

537 uPro.....HisArgProArgCysHisArgPro 546
||| ||| |||::: |||
222 GCCGAACGCAACGACATGCCATGCATCATTAACCA 186

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-4

seq_documentation_block:
/ Sequence 4, Application US/08864038A
/ Patent No. 6001592
/ GENERAL INFORMATION:
/ APPLICANT: Kunio NAKASHIMA et al.
/ TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
/ TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
/ TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
/ TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
/ TITLE OF INVENTION: TO SAID POLYPEPTIDE
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: 812-5 Hirano
/ STREET: Ieshinden
/ CITY: Tsu-city
/ STATE: Mie-prefecture
/ COUNTRY: JAPAN
/ ZIP: 514-01
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Microsoft Windows 95
/ SOFTWARE: Word Perfect 6.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US//08/864,038A
/ FILING DATE: May 28, 1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 8-184459
/ FILING DATE: 15-July-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: C. Bruce Hamburg
/ REGISTRATION NUMBER: 22,389
/ REFERENCE/DOCKET NUMBER: F-5610
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)986-2340
/ TELEFAX: (212)953-7733
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 331
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA to mRNA
/ ORIGINAL SOURCE:
/ ORGANISM: Pinclada fucata
/ CELL TYPE: mantle epithelial cell
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: from 50 to 2263
/ IDENTIFICATION METHOD: P (by similarity to some other pattern)
/ US-08-864-038A-4

```

```
alignment_scores:
  Quality: 381.50
  Ratio: 1.408
  Length: 597
  Gaps: 25
```

Percent Similarity: 45.394 Percent Identity: 25.628

Alignment block:

US-09-668-119-3 x US-08-864-038A-4/rev ..

Align seg 1/1 to reverse of: US-08-864-038A-4 from: 1 to: 3331

```

50 AlaSerValSerAspPro.....MetAsnAlaLeuGI 60
   |||.....|
1680 GCGGACGACGCGGATTCACCTCCACGACGATTTGCTTCTTCAAGCCCTTCT 1631
60 nSerLeuThrGlyLPro.....AlaAlaGlyAlaA 71
   |||.....|
1630 AAGTCTCTTCCTCCACCACTGATGACGACGCGGACGCGCTGACGACG 1581
71 laGlylleGlyMetProProArGlyProGlyGlnSerLeuGlyGlyMet 87
   |||.....|
1580 CTGCGGACGATGCCCTACCAAGACGCGCT..... 1552
88 GlySerPheGlyAlaMetGlyGlnProMetSerLeuSerGlyGlnProPr 104
   |||.....|
1551 .....AGTCTCCAAAGTCCACCAAAACCTCTCC 1523
104 oProGlyThrSerGlyMetAlaProHisSerMetAlaValAlaSerThra 121
   |||.....|
1522 ACCT.....GCACCGACGACGACGTAAGTGCACGCGCTAAAG 1488
121 lArhrPro.GlnThrGlnLeuGlnLeuGlnInValAlaLeuGlnGlnI 137
   |||.....|
1487 CTCCGCTCCGCTCCACCTCAAGTCCACCTCCGCTCCGCTCCGACGCG 1458
137 nGlnGlnGlnGlnGlnPheGlnGlnGln..... 147
   |||.....|
1437 AGAGCAGCAGCTAGACtCTCAGCTCCACCGCGCTCTCTCCGTTACC 1388
148 .....GlnAlaAlaLeuGlnGlnGln 154
   |||.....|
1387 ACCTCCGTTACCAACGACGCTCCACCTCCACGCTGAAGCTGATGACATG 1338
155 GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGln..... 166
   |||.....|
1337 CAGATGCGAAGCAGATTCCTGAGCTTTCAATAGACTGCAGATCCCTTT 1288
167 .GlnSerAlaMetGlnGlnGlnPheGlnAlaValVal..... 178
   |||.....|
1287 AAGATCTCTTCAAAAGCTCAATTGAGATTGTGCTTGTCTACTGCTGA 1238
178 ..... 178
1237 TGCTTTAGCAGAGGCTTTTGACAGAGGCTCCATTCCTGAGAGGTCTT 1188
178 ..... 178
1187 TAATGCTCTTAATGCTGCCCTTATACCACTATTGTCTAGAGCTGAA 1138
179 .GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGln..... 192
   |||.....|
1137 GCAGATGACGACGCGCTGACGCGGACGCGGACGACGACGCGCTCCGCC 1088
193 .....HisLeuLeuLeuHisHisGlnGlnGln 203
   |||.....|
1087 TCCCTCCCAACCACTCCACTCTCTCCGCCACCAACCAAGCGGCTGCGGG 1038
204 GlnlleGlnGlnGlnGlnGln.....LeuGI 213
   |||.....|
1037 CAGCTGACGACGCGGACGACGACGCTGCGGCGCTCTCTTCTTCGGA 988
213 nArGlyAlaGlnLeuGlnLeuGlnGlnGln..... 224
   |||.....|
987 CCGGCTTACTCTCTCTCCACCTAGAAACCAACACCGAGTCTCC 938
225 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeu 238
   |||.....|

```

```

937 ACCTCCGACGACGCGGCGGACGACGACGCGGACGACGATCTCC 888
239 GlnAlaGlnProIleGlnGlnProIleGlnGlnProIleProPr 255
   |||.....|
887 CATAT.....CTCCGAGGCGCACCAAGACCTCCGAGGCTCCCAAGTCCGCC 842
255 oProSerGlnAlaLeuProGlnGlnGlnGlnGlnGlnGlnHisThrGlnH 272
   |||.....|
841 ACCGAGTCTCCA..... 829
272 lshIsGlnProProGlnProGlnGlnProProValAlaGlnGln 288
   |||.....|
828 .....AGTCCGCCACCGAGTCTCCCAAGTCCACCTCCG..... 796
289 ProSerGlnLeuProProGlnGlnGlnGlnGlnGlnGlnGlnValSerGlnAl 305
   |||.....|
795 .....CTCTCTCAGCGGACGCA.....GCGGCTCCGCC 767
305 aglnAlaLeuProGlyGlnMetLeuTyThrGlnProProLeuLysPheV 322
   |||.....|
766 TGCAGCAGCAGCTCTCCAAAGTCTACTGCGCTCTCTCTCCGACGCGCG 717
322 aLArgAlaProMetValValGlnGlnProProValGlnProGlnGln 338
   |||.....|
716 CACGCGCTGCGGCTCCGCGGACGACGCTCCACCACTCCACCCCGCG... 670
339 GlnGlnGlnThrAlaValGlnThrAlaGlnAlaAlaGlnMetValAlaPr 355
   |||.....|
669 .....GCACGACGCGGACGCTCCGACGACGACGACGACGACGCTCC 626
355 oglyValGlnValSerGlnSerSerLeuProMetLeuSerSerProSerP 372
   |||.....|
625 ACCAGCTCCACGAGGCGCAAGAGCTGCACCAAGA...TCATGAAATCTA 579
372 roGlyGlnGlnAlaGlnThrProGlnSerMetPro...ProProGln 387
   |||.....|
578 AATCAATTAATGCTCTCCACCGCCCAAGTCCGCCAAGACTCTCAAGTCCG 529
388 ProSerProGlnProGlyGlnProSerSerGlnProAsnSerAsnValSe 404
   |||.....|
528 CCAGTCCACCTCCGACCTCCG..... 505
404 rSerGlyProAlaProSerProSerSerPheLeuProSerProSerProG 421
   |||.....|
504 .....CCCAATCCAAATCTCCAGT.....CCAGACCTGCTCTCG 471
421 InProSerGlnSerProValThrAlaArgThrProGlnAsnPheSerVal 437
   |||.....|
470 CTCTGCGCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 421
438 ProSerProGlyProLeuAsnThrProValAsnProSerSerValMetSe 454
   |||.....|
420 CCAGCAGCAGCTCCGCGGCGGACGCTCCCATTCACGCTCATCTCCAGT 371
454 rProAlaGlySerSer.GlnAlaGlnGlnGlnGlnGlnGlnGlnGlnLys 470
   |||.....|
370 CCATTCATATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 321
471 LysGlnLeuSerLysTyrlleGlnProLeuArgArgMetlleAsnLysII 487
   |||.....|
320 CACCGC.....GCGCCCGCGCTCCG..... 300
487 eaSpLysAsnGlnAspArgLysLysAspLeuSerLysMetLysSerLeuII 504
   |||.....|
299 .....GTCGCGCTCC 290
504 euAspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
   |||.....|
289 TCCAGACCGCAGCTCCCAAGCGCTCCGCGCTCCGACAC..... 246
521 LysCysGlnIleAlaLeuGlnLysLeuLysAsnAspMetArgCysProLe 537
   |||.....|
245 .....GCTAAAGACCTGCACCTGCTCAT 223

```


Fri Mar 1 09:12:02 2002

```

? TELEPHONE: (212) 278-0400
? TELEFAX: (212) 391-0525
? INFORMATION FOR SEQ ID NO: 20:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 32207 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-757-669A-20

```

Quality:	314.00	Length:	343
Ratio:	1.602	Gaps:	9
Percent Similarity:	57.143	Percent Identity:	30.904

alignment_block:
US-09-6668-119-3 x US-08-757-669A-20/rev

Align seg 1/1 to reverse of: US-08-757-669A-20 from: 1 to: 32207

```

7  AlaHisSerLysSerSerLysAspMetGluSerHisValPheLeuLysAl 23
   ::|||:::  ||  ::  ::|||  ::
20489 AGCCACAAACAGCAGGAGCCACACAGCAGGAGCCACACAGCAGCAGAGCCA 20440

```

```

23  alyshirargaspgl jTyrLeuSerLeuValAlaargLeuIleHis 40
      :::::||||:
20439 CAGCAGCAGGAGC.....CACA 20423

```

40 heargspIleHisa3nLysIysSerGlnAlaSerValSer.aspProme 56
 |||:::|||||:::||||| |||
 20422 GCAGCAGGAGCCACA3CAGCAGAGAGCCACACAGCAGCAGAGAGCCACAGCACC 20373

56 taaalaleugInsectleuthrglyglyProalaalaglYalaalaclyI 73
::||| :: ::|||:: |||||::|:::
20372 AGGAGCC....ACtgcagcagagaccacAGcAGCGGGAGccCACgCAG 20329

73 IeGlymetProProArGgLyProGlyGlnSerLeuGlyGlyMetGlySer 89
 ||| ||| :: ||| ::||| ::
 20328 CGGGAAGCCCCA....GCACCGCGGAGCCACAGCAGCGGGAGCCACAGCA 20285

```

90 PheGlyAlaMetGly31nPrometSerLeuSerglyInProProG1 106
      |||||      ::      ::|||      ||
20284 GCGGGAGCCACAGCA3CGGGAGCCACAGCAGCGGAGCCACAGCAGCGGG 20235

```

```

106 yThrSerGlyMetAlaProHisSerMetAlaValSerThrAlaThr. 122
|:::| |:::|
20234 AGCCACA.....GCAGCGGAGGCACACAGCAGCAG 20206

```

123 ProGlnThrGlnLeu;IlnLeuGlnIlnValAlaLeuGlnGlnGlnGlnI 139
::: ||| ::| ||||| | |||||:::|
20205 GATGAGCAGCAGCAGIATGAGCAGCAGCAGATGAGCAGCAGCAGCATGA 20156

139 nGInglngInPheGlngInglngInglnglnAlaAlaLeuGlngInglng 156
:::|||||||:::|||||||:::|||||||:::|||||||:::|||||||
20155 GCAGCAGCAGCATGAACAGCAAGCAGCATGAGCAGCAGCAGCATTCACCC 20106

[illegible]

173 GlmpheglAlaValValGInglnglInglnglInleuglnglInglngl 189
::: ||| ::||| |||::: | ||| |||:::
GTCCTCGTGGAGCAATGACGTTGCTGCTGCTGCTGCTGCTGCTGCT

```

189 nglnglInHlslLeuIeLysLeuHlshlsglInasnglInglInIle 206
    :|||||||      ::      :::::|||||||      :

```

206 InGIInGIInGInGInleugInargIleAlaInleugInleugInGIn 222
:::|||||::: : |||::: ||| ::|||

19961 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAG 19912

us-09-668-119-3.rni

Page 10

```

223 glnnglnglnglnglnglnglnglnglnglnglnglnglnglnglnglnalalaugl 239
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
19911 CAGCAGAGATGACACAGACACAGCAGATGACAGCAGCAGCAGATGACGA 19862
239 ualnglnproprolleglnglnproprometglnlnglnprolnproprop 256
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
19861 GGAGCAGCAGATGACGACGACGACGAGATGACGACGACGACGATGAC 19812
256 roserglnalaleuproglnleugln.....glnmethishisthr 270
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
19811 ACCGACGACGAGATGACGACGACGACGAGATGACGACGACGACGAT 19762
271 glnhishistglnproproprolnprolnglnlnproprovalalaglnas 287
:::|||||:::|||||:::|||||:::|||||:::|||||:::
19761 GAGCAGCAGCAGCAGATGAAACGACGACGACGACGACGACGAGCAGCA 19712
287 nlnproserglnleuproproglnglnserglnthglnproleuvalserg 304
:::|||||:::|||||:::|||||:::|||||:::|||||:::
19711 GGAGAGCAGCAGCGAGGAGTTAGAGGACGACGACGACGAGATTAGAGATC 19662
304 lnlalglalaleuproglnmetleutyrtlnrglnproproleuyls 320
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
19661 AGCAGCAGGATTTAGAGGACGAG.....GAGCAGGATTTAGAG 19624
321 phevalarglnalaprometvalvalglnlnglnproprovalglnproglng 337
:::|||||:::|||||:::|||||:::|||||:::|||||:::
19623 GAGCAGCAGCAGCAGATTGAGGACGACGACGACGACGAGATTAGCAGCAGA 19574
337 lglnglnglnglnlnlralalvalgln 345
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
19573 GCAGGAGTTAGGAGCAGCAGCAGCAG 19549

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-991-300-1

seq_documentation_block:
Sequence 1, Application US/08991300
Patent No. 5973225
GENERAL INFORMATION:
APPLICANT: D'OVIDIO, RENATO
APPLICANT: PORCEDDU, ENRICO
APPLICANT: MERCHITELLI, CINZIA
APPLICANT: CARRELLI, LUISA ERCOLI
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENES
TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,300
FILING DATE: 16-DEC-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IT MI 96/A 002663
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2264-0201-0X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220

```


Fri Mar 1 09:12:02 2002

us-09-668-119-3.rni

Page 14

;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5318 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-09-668-119-3

alignment_scores:
Quality: 309.50 Length: 232
Ratio: 2.437 Gaps: 9
Percent Similarity: 54.741 Percent Identity: 43.966

alignment_block:
US-09-668-119-3 x US-08-928-361B-3 ..

Align seg 1/1 to: US-08-928-361B-3 from: 1 to: 5318

```
113 HsSerMetAlaValAlaSerThrAlaThrProGlnThrGlnLeuGlnLe 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 CATACACTAAATGTGTGGAGTGAACACACACACACACACACACACT 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
129 uGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGlnPheGlnGln 146
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 ACTACTACTACTACTCTCGACACACACACACACACACACACACACAC 644
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 lnglngln.....AlaAlaLeuGlnGlnGlnGlnGlnGlnGln 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 694
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 GlnGlnGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
695 CAACACACACACACACACACACACACACACACACACACACACTACT 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 .....PheGlnAlaValAlaGlnGlnGlnGlnGlnGlnGln 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
745 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 794
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 lnglnglnglnglnglnglnglnglnglnglnglnglnglnglng 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
795 AACACACACTACTACTACTACTACTACTACTACTACTACTACTACT 829
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 CAACACACACACACACACACTACTACTACTACTACTACTACTACT 876
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 uGlnLeuGlnGlnGlnGln.....GlnGlnGln 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
877 ACAACCCACACACACACACACACACACACACACTACTACTACTACT 926
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 lnglnglnglnglnglnglnglnglnglnglnglnglnglnglng 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
927 AACCACTACTACTACTACTACTACTACTACTACTACTACTACTACT 976
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 lIeGlnGlnProMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
977 ...CAACACACACACACACACACTACTACTACTACTACTACTACT 1023
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 .....AlaLeuProGlnGlnL 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1024 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1073
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 euGlnGlnMetLHisLSThrGlnHisLSThrProGlnProGlnProGln 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1074 AACCACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 GlnProGlnValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1124 CAACCTACACGAAACCTACACACACACACACACTACTACTACTACT 1173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 oProGlnSerGlnThrGlnProLeuValSerGlnAlaGlnAlaLeu 308
```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1174 ACCAAGAACACACACACACTACTACTACTACTACTACTACTACTACT 1219

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-928-361B-2
seq_documentation_block:
Sequence 2, Application US/08928361B
Patent No. 6071518

GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verry, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-2

alignment_scores:
Quality: 296.50 Length: 246
Ratio: 2.335 Gaps: 13
Percent Similarity: 51.626 Percent Identity: 41.057

alignment_block:
US-09-668-119-3 x US-08-928-361B-2 ..

Align seg 1/1 to: US-08-928-361B-2 from: 1 to: 5511

```
113 HsSerMetAlaValAlaSerThrAlaThrProGlnThrGlnLeuGlnLe 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
896 CATACACTAAATGTGTGGAGTGAACACACACACACACACACACT 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
129 uGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnPheGlnGln 145
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 995
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 .....GlnGlnGlnAlaValAlaLeuGln 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 TACTACACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1045
```

```

153 GlnGlnGlnGlnGln.....GlnGlnGlnGlnGlnPheGlnAl 165
1046 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1095
165 aglnGlnSerAlaMetGlnGlnGlnPheGlnAlaValGlnGlnGln 182
1096 AGCACA.....CTACAACAACAACAACAACAACAACAACAACA 1115
182 GlnGlnLeu.....GlnGlnGlnGlnGlnGlnGlnGlnHisLeu 195
1116 CACAACTACCAAGAAACAACAACAACAACAACAACAACAACAACA 1165
196 LysLeuHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 212
1166 .....CTACAACAACAACAACAACAACAACAACAACA 1188
212 GlnArgIleAlaGlnLeuGlnLeu.....GlnG 222
1189 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 1238
222 GlnGlnGlnGln.....GlnGlnGlnGlnGlnGlnGlnGln 233
1239 AACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 1288
234 GlnGlnGlnAlaLeuGlnAlaGlnPropGlnGlnGlnGlnGlnGln 250
1289 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1332
250 GlnPropGlnPro.....ProProSerGln.....A 259
1333 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1382
259 LysLeuPropGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 273
1383 TACTACCAACAACAACAACAACAACAACAACAACAACAACAACA 1432
274 .....GlnPropProGlnPropGln.....GlnPropProValAl 285
1433 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 1482
285 aglnGlnGlnPropSerGln.....LeuPropGlnGln 296
1483 AAGAAACAACAACAACAACAACAACAACAACAACAACAACAACA 1532
296 erglnGlnGlnPropLeuValSerGlnAlaGlnAlaLeu 308
1533 ACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 1570

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-928-361B-1

seq_documentation_block:
Sequence 1, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIRSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7334 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-1

```

```

alignment_scores:
Quality: 296.50 Length: 246
Ratio: 2.335 Gaps: 13
Percent Similarity: 51.626 Percent Identity: 41.057

```

alignment_block:

US-09-668-119-3 x US-08-928-361B-1 ..

Align seg 1/1 to: US-08-928-361B-1 from: 1 to: 7334

```

113 HisSerMetAlaValAlaSerThrAlaThrProGlnThrGlnLeuGln 129
2564 CATACACTAATGTGTGGAGTGAACACACACACACACACACACAC 2613
129 uGlnGlnAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 145
2614 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 2663
146 .....GlnGlnGlnAlaAlaLeuGln 152
2664 TACTACACTACTACTACTACTACTACTACTACTACTACTACTACTAC 2713
153 GlnGlnGlnGlnGln.....GlnGlnGlnGlnGlnPheGlnAl 165
2714 CACAACAACAACAACAACAACAACAACAACAACAACAACAACA 2763
165 aglnGlnSerAlaMetGlnGlnGlnGlnPheGlnAlaValGlnGlnGln 182
2764 AGCACA.....CTACAACAACAACAACAACAACAACAACAACA 2783
182 GlnGlnLeu.....GlnGlnGlnGlnGlnGlnGlnGlnHisLeu 195
2784 CACAACCTACCAAGAAACAACAACAACAACAACAACAACAACAACA 2833
196 LysLeuHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 212
2834 .....CTACAACAACAACAACAACAACAACAACAACA 2856
212 uGlnArgIleAlaGlnLeuGlnLeu.....GlnG 222
2857 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 2906
222 GlnGlnGlnGln.....GlnGlnGlnGlnGlnGlnGlnGln 233
2907 AACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 2956
234 GlnGlnGlnAlaLeuGlnAlaGlnPropGlnGlnGlnGlnGlnGln 250
2957 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 3000

```


Db 609 PPM^{SV}ATVTRGPPPT-----RPASVIAPSSSV--PKSQPR--PSP^{TR}TRPPTRATIR 657

Db 844 FSNMNNANFLNCPRGPGQGNAPGNMPPQ0000PQ0000PPRGPGQSNPNAPGGMANN 903
 QY 215 IADLQ0000000000000000ALEP----- 240
 Db 904 ATQ00000000000000000000ATTTTLMKQTQDLHISQGGSHGIVSAGQLH 963
 QY 241 -----OPPIQPPMOOPPPPSQALPQOLQOMHNTQH-----H 273
 Db 964 LSSDMKSNVSYAAGCVFFSSQ00AA00000000PQGNAPRNQ0000QPHGGNAGANGG 1023
 QY 274 QPPPGQPP-----PVAQNPQSLPPQSQTQPLVSAQALPGLMLYTQPL 319
 Db 1024 PNPQ0000PQNMNNNSVPSDFSLQSQSMNFTQ00000000AAAAA0000QA----- 1079
 QY 320 KFRAMVNVQPPPVQVQ00Q0TAVQTAQAQNVARGVQSSSLPMLSPSPG----- 373
 Db 1080 ----AA00000000PPMQRQRTQAGAAAAA-----AAQA0AAANANGPGGNVPLM 1131
 QY 374 -QQVTPQSMPPPPQPPQPPSQPNSVSGPAPSPSFLPSPQPPQSPYTARTP 432
 Db 1132 Q0000TP-----GGVYVAGSGNASYG-----VPVSAGGP 1161
 QY 433 QNFS-----VPSGPLNTPVPSVSPAGSSQAE 462
 Db 1163 NNGAMNQLGGPMGMDPMQMGPGVPIPNMQMNPNGAPNAQ 1204

RESULT 12

T30160
 hypothetical protein C37A2.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T30160
 R:Le, T.T., Kemp, K.; Scheet, P.
 Submitted to the EMBL Data Library, April 1997
 A:Description: The sequence of C. elegans cosmid C37A2.
 A:Reference number: Z20746
 A:Accession: T30160
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1234 <LEFT>
 A:Cross-references: EMBL:U97194; PIDN:AAB52447.1; GSPDB:GN00019; CESP:C37A2.2
 A:Experimental source: strain Bristol N2; clone C37A2
 A:Genetics:
 A:Gene: CESP:C37A2.2
 A:Map position: 1
 A:Intons: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

Query Match 12.4%; Score 372; DB 2; Length 1234;
 Best Local Similarity 25.4%; Pred. No. 1.2e-09;
 Matches 170; Conservative 50; Mismatches 169; Indels 280; Gaps 29;
 QY 71 AGIGMPRGPG---QSL---GMSGFGAMGQPMSLSGQPPGTSGMAPHMAVSTAN 122
 Db 403 SSGSGPRTVVGTRPQGITPQPSLQPMGSLGPRTA-----PRGSGPMNP----- 447
 QY 123 PQTQLQLQVVALQ000000FQ0000A-----LQ00000000QF---QAQ0 167
 Db 448 -----QQRIRQ0000AAPSASNSPLLNLNSNQPPQ000YMPGPSA0G 491
 QY 168 SAMQ0QFQAVVQ000LQ00000000HILKHNQ0Q0I00000QLRIAGLQ000000 227
 Db 492 LSMQ0--IAAIQ0000-----HQYQQRILQ0000-----QAMMQ00 526
 QY 228 Q00000000A-----LEAQP-----IQ0PPM----- 249
 Db 527 QQQVQ000QAAPRTTMRGHRGPFRTNAGRGRTGPRIRYAMAGPRMYHAGPQGM 586
 QY 250 -----Q0P0PPSA---LPQ0LQ0MHTH0RPPR00P0PVAQ0NP0SLP- 294
 Db 587 IQRNYSYRGAQ0FRPPQ0SQ0P0P0Q000---QPRAPQ0LQ0P0VQDTSAVAEPK 642

QY 295 -----Q0TQPLVSAQALPQ0MLYTQPLKFRVAP-----MVYQ0P- 331
 Db 643 KKKRTTKQKRAAAAALAEQ000000Q0MAVYQ000000000000000000000000 702
 QY 332 -----PVQPVQ00Q0TAVQTAQA---QNVAPGVQS 362
 Db 703 YPGQYGMPPPGCAFPPGYLPGAQPTQ0QIMQ0000LQ00RMAQM0000AQ0CGCP 762
 QY 363 SLPM-----LSSPSGQ0VQTPQSM-----PP-----P 386
 Db 763 Q0PMQ0MPQ0RQ0LPVYPPRGNNSTPDVAVN000NP1PGATMQHRLSGEPAPPVSSGKVP 822
 QY 387 QPSPPQ0PSSQ0PNNSVSGPAPSPSFLPSPQPPQSPPTARTPQNFVSPSGPLNTP 446
 Db 823 Q0000GSRHRDSSASVSG-----SHTPGQ0GPPQSGSEPTA-----VP-PGPPQN- 867
 QY 447 VNPSSVSPAGSSQAEQ0YLDKQLK0SKYTEPLRRM-----INKTD----- 488
 Db 868 -----NPGSGSDIGEKAIVDQILNLS---EP1ADLDGLDGLDLEPMDVQDQTPS 917
 QY 489 -KNEDKKDLKMSKSLDILTDPSKRCPLKTLQKCE1ALEKNDMRCPLP----- 538
 Db 918 TSGERNERSDRLASTELVDEVAAGRAGAPNAELAIIGRRGSVPIPLPASPGRNL 977
 QY 539 ---HRRPCH 544
 Db 978 SHEHNMRAH 986

RESULT 13

T13606
 hypothetical protein 87B1.5 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13606; S23632
 R:Murphy, L.; Harris, D.; Bartell, B.
 Submitted to the EMBL Data Library, April 1999
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: Z17668
 A:Accession: T13606
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1589 <MUR>
 A:Cross-references: EMBL:Z96269; NID:el355202; PID:el251078; PIDN:CAB10975.1
 R:Decamillis, M.; Cheng, N.; Plerre, D.; Brock, H.W.
 Genes Dev. 6, 223-232, 1992
 A:Title: The polyhomeotic gene of Drosophila encodes a chromatin protein that shares
 A:Reference number: S23632; MUID:92146957
 A:Accession: S23632
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1010, 'V', 1012-1192, 'L', 1194-1274, 'T', 1276-1589 <DEC>
 A:Cross-references: EMBL:X63672; NID:g11056; PIDN:CAAA5211.1; PID:g11057
 C:Genetics:
 A:Gene: FlyBase:ph-p
 A:Cross-references: FlyBase:FBgn0004861; FlyBase:FBgn0004860
 A:Intons: 12/2; 595/1; 745/2; 1340/1
 C:Superfamily: SAM homology
 C:Keywords: DNA binding; nucleus
 F:74-80, 247-285, 411-450, 494-650, 727-737, 775-955, 1032-1061/Region: glutamine-rich
 F:1510-1576/Domain: SAM homology <SAM>

Query Match 12.1%; Score 365; DB 2; Length 1589;
 Best Local Similarity 29.7%; Pred. No. 3e-09;
 Matches 158; Conservative 57; Mismatches 177; Indels 140; Gaps 24;
 QY 7 ASKSKSDMESHVFLKATRDY-----LSVA---RLIHFDHNHKKQA 50
 Db 674 ASSVSTQTAQ0NSLKKAKRNQ0PVPALATLKTEIGQVAGQNKVVGH1TTVQ00Q0AT 733
 QY 51 SVSDPMNAL0SLTGGPAGAGACIGMPRPQ0SLGMSGFGAMGQPMSLSGQPPGTSGM 110

Fri Mar 1 09:12:03 2002

us-09-668-119-3.rpt



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 19:37:53 ; Search time 16.91 Seconds
(without alignments)
1255.407 Million cell updates/sec

Title: US-09-668-119-3
Perfect score: 3010
Sequence: 1 MRKAGVAHSSSSKDMESHVF.....WPTSAHLSSTIPCTAHSPFP 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	383	12.7	708	1	GBF_DICDI
2	383	12.7	966	1	SSN6_YEAST
3	378	12.6	1596	1	MAM_DROME
4	366	12.2	1589	1	PHP_DROME
5	327.5	10.9	1008	1	GALY_KLUTA
6	327.5	10.9	2038	1	FSH_DROME
7	326.5	10.8	905	1	SNF5_YEAST
8	310	10.3	255	1	LP61_EIMTE
9	310	10.3	2124	1	Y192_HUMAN
10	308	10.2	356	1	GLTA_WHEAT
11	298.5	9.9	838	1	GLT4_WHEAT
12	298	9.9	1403	1	PRO_DROME
13	296.5	9.9	648	1	KAPC_DICDI
14	296	9.8	839	1	GLT5_WHEAT
15	295.5	9.8	1123	1	DC11_DROME
16	294	9.8	519	1	ELAY_DROVI
17	294	9.8	2442	1	CBP_HUMAN
18	292	9.7	1589	1	DC13_DROME
19	290.5	9.7	738	1	YKF4_YEAST
20	285	9.5	467	1	INVO_MOUSE
21	284.5	9.5	1231	1	YK73_CAEL
22	284.5	9.5	1902	1	SMF1_HUMAN
23	283	9.4	313	1	GDA7_WHEAT
24	283	9.4	319	1	GDA5_WHEAT
25	283	9.4	2441	1	CBP_MOUSE
26	282	9.4	848	1	YBVB_YEAST
27	282	9.4	1638	1	BRM_DROME
28	279.5	9.3	493	1	HMBB_DROME
29	278.5	9.3	550	1	CCR_DROME
30	278.5	9.3	660	1	GLT3_WHEAT
31	278.5	9.3	950	1	DC12_DROME
32	274	9.1	648	1	GLT0_WHEAT
33	273.5	9.1	3828	1	TRX_DROVI

34	273	9.1	1028	1	OVO_DROME	P51521 drosophila
35	272.5	9.1	618	1	ZEST_DROVI	Q24762 drosophila
36	271.5	9.0	347	1	INVO_PIG	P18175 sus scrofa
37	270.5	9.0	395	1	SRY_MOUSE	O05738 mus musculus
38	269.5	9.0	1331	1	NFT3_HUMAN	O94916 homo sapien
39	267.5	8.9	700	1	BIB_DROME	P23645 drosophila
40	264	8.8	468	1	YOC1_CAEL	O09260 caenorhabdi
41	264	8.8	1023	1	CLOC_DROME	O61735 drosophila
42	264	8.8	1443	1	E75C_DROME	P13055 drosophila
43	261.5	8.7	1081	1	GALY_YEAST	P19659 saccharomyc
44	256.5	8.5	1556	1	PRO_DROVI	O09651 drosophila
45	256	8.5	2414	1	P300_HUMAN	O09472 homo sapien

ALIGNMENTS

RESULT	ID	GBF_DICDI	STANDARD	PRT	708 AA.
AC	P36417	GBF_DICDI			
DT	01-JUN-1994	(Rel. 29, Created)			
DT	01-JUN-1994	(Rel. 29, Last sequence update)			
DT	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	G-BOX BINDING FACTOR (GBF).				
CN	GBFA.				
OS	Dictyostelium discoideum (Slime mold).				
OC	Eukaryota; Mycelozoa; Dictyostellida; Dictyostelium.				
ON	NCBI_Taxid=44689;				
RX	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	SPRAIN=AX3;				
RA	MEDLINE=94170994; PubMed=8125261;				
RT	Schmitzler G.R., Fischer W.H., Firtel R.A.;				
RT	"Cloning and characterization of the G-box binding factor, an				
RT	essential component of the developmental switch between early and				
RT	late development in Dictyostelium."				
RL	Genes Dev. 8:502-514(1994).				
CC	-!- FUNCTION: CAMP-RESPONSIVE TRANSCRIPTIONAL ACTIVATOR REGULATING				
CC	LATE GENE EXPRESSION. ESSENTIAL COMPONENT OF THE DEVELOPMENTAL				
CC	SWITCH BETWEEN EARLY AND LATE DEVELOPMENT. BINDS TO A NUMBER OF				
CC	CA/GT-RICH GENE REGULATORY ELEMENTS.				
CC	-!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC					
DR	EMBL; L29075; AAA21021.1; -.				
DR	TRANSFAC; T00315; -.				
DR	Dictydb; DD02046; gbfa.				
KW	Transcription regulation; Activator; DNA-binding; Nuclear protein;				
KW	Repeat.				
FT	REPEAT	339	368	1.	
FT	REPEAT	481	510	2.	
FT	DOMAIN	11	21	POLY-SER.	
FT	DOMAIN	115	263	GLN-RICH.	
FT	DOMAIN	270	292	POLY-RICH.	
FT	DOMAIN	549	557	POLY-ASN.	
FT	DOMAIN	708	79268	MW; 84B6B0F4FACACCA CRC64;	
SO	SEQUENCE				
Qy	44	HNKSQASVSPMNALOSITGPPAAGAGTGMPPRG-----PGQS-LGG 86			
Db	6	HHQGNSSSSSSSPQITGGSDLSNITSALPLPUPSIFTTAQNQMNPILFPPTSLGG 65			

Query Match 12.7%; Score 383; DB 1; Length 708;

Best Local Similarity 24.4%; Pred. No. 1.5e-09;

Matches 170; Conservative 57; Mismatches 200; Indels 270; Gaps 23;


```

OY 189 QQQQHILIKLHNOQQOIQQQQOQLORIALQOQQQQQQQQQOQOQOQALENPPIOQPP 248
DB 567 QQQQQQ--QOQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 624
OY 249 MOOP-----OPPPSOALP---QOQOQHHPHHPPOPPPOOP--PVAONOPSO--LPPOS 296
DB 625 QLOPPSTQAMHPQSSQLPPOQOQLOLQVQHPQOQLOGQOQAPLPFLQHNEQNVLPQK 684
OY 297 QTPPLV-SQAQALPQOALTPPLKFNAP--MVVQOPP-----VOPQVQOQQTAVQTA 347
DB 685 YMEGAHHTLVDAVSSSTHTENTKSPQPTHAIPQAATGITNAEPQVKQKLNPSNS 744
OY 348 QAAQNVAPCVQVSQSLPMLSSPSPG-----QOQVTP--OSMP--PPQSPQPG 393
DB 745 NINKLVNTATSIENAKSEVSNQSPAVVESNTNTNTSOEKPVPKANSIPVIGAOEPPQEA 804
OY 394 OPSQOPNSNVSSGPAPSPSPFLP-SPSPQSPQVPTARTPON-----FSVP-- 438
DB 805 SPPEATKAASVSPSTKPLNTEPESSVQPTVSSSTTKANDOSTAETIELSTAVPAE 864
OY 439 -SP-----GPLNTPVNPSSVMSPPASSQAOEQ-----YLD 468
DB 865 ASPVEDEVQHSKEENGTEASAPSTEEAPASDAEKQODEFATTTIVIKPTLEME 924
OY 469 KIKOLSKYTE-----PLRMINKIDKNE 492
DB 925 TYKEAKMREBOTSOEKSPOENTLPRENVVQVEEDEN 963

RESULT 3
MAM_DROME STANDARD: PRT; 1596 AA.
AC 1 P21519;
DB 01-MAY-1991 (Rel. 18, Created)
DB 01-MAY-1991 (Rel. 18, Last sequence update)
DB 01-MAR-1992 (Rel. 21, Last annotation update)
DE NEUROGENIC PROTEIN MASTERMIND.
GN MAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE=91065516; PubMed=1701150;
RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA Yedvobnick B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers.";
RL Genes Dev. 4:1688-1700(1990).
CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DB EMBL: X54251; CAA38152.1; -.
DB PIR: A33106; A33106.
DB PIR: A36391; A36391.
DB FlyBase: FBgn0002643; mam.
KW Neurogenesis; Nuclear protein; Repeat.
FT DOMAIN 20 84 ARG/LYS-RICH (BASIC).
FT DOMAIN 127 190 GLN-RICH.
FT DOMAIN 196 219 GLN-RICH.
FT DOMAIN 259 304 ASN-RICH.
FT DOMAIN 355 388 GLY/ASN-RICH.
FT DOMAIN 392 406 GLN-RICH.
FT DOMAIN 407 440 GLY-RICH.
FT DOMAIN 651 671 GLN-RICH.
FT DOMAIN 700 714 GLN-RICH.
FT DOMAIN 759 816 GLN-RICH.
FT DOMAIN 987 996 ALA-RICH.
FT DOMAIN 1060 1079 5 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1092 1107 8 X 2 AA TANDEM REPEATS OF V-G.
FT DOMAIN 1237 1252 7 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1492 1496 POLY-THR.
FT DOMAIN 1559 1592 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 1596 AA; 167717 MW; B944D86EF359D605 CRC64;

Query Match 12.68; Score 378; DB 1; Length 1596;
Best Local Similarity 30.28; Pred. No. 4; Ce-09;
Matches 157; Conservative 30; Mismatches 143; Indels 190; Gaps 24;

OY 46 KKSQASVSDPMNALQ-SLTGSPAGAGIGMPRPGQSLGMSGFGAMGQPMSLSGOP- 103
DB 537 KVEPPNODLINSLNKSEGLGHGFGGLDLPGMKMG-----GNPQGGSGFPN 589
OY 104 -PPTSGMAPH-----SMAVSTAPQTOLOLQVALQ----- 135
DB 590 GPNGTGAGAPNAGNGSGNLMSEHPLAOTLKQMAEHOHKNMGMGGFRRPRGHN 649
OY 136 -QOQQQQQFOQOQOQAAQOQ-----QOQQQQQOQF 163
DB 650 PQOQQQQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 709
OY 164 QAOQSAM-----QOQFQAVVQ-----QOQOQOQOQOQ 190
DB 710 QPQOQHLPQFHQKGRPRGAGMNVQNFIDIKQLFLFSSQNDPDKRLQOQOQAMQOQOQ 769
OY 191 QOHLIKLHNOQOQIQ-----QOQOQLOQRIALQOQOQOQOQOQOQOQOQOQOQOQ 241
DB 770 QO-----HHQOQOPRMSGVPRNFKNQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 824
OY 242 -----PP-----IQP-----PMQOPRPPSQALPQOQLOQMHT--QHQP----- 275
DB 825 PMAAANFLNCPRGSPGNGNOQPGNLAQOQOQ-----GAGPQOQOQOQOQOQOQOQOQOQ 881
OY 276 -----PPQOPFPVAVQNPQSOQLPQSOQOPL-VSQ-----AQAIPQMLTQPLKFNVR 323
DB 882 GGNFPMNAPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 939
OY 324 APNVVQOPPV-QPQVQOQQTAVQTAQAQNVAPCVQVSQSLPMLSSPSPQOQVQTPQS 381
DB 940 VSAVAQOQVFPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 981
OY 382 MPP-----PPQSPQOPGSPQSPNSVSGPAPS 410
DB 982 GNAGGQVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV 1021

RESULT 4
PBP_DROME STANDARD: PRT; 1589 AA.
AC P39769;
DB 01-FEB-1995 (Rel. 31, Created)
DB 01-FEB-1995 (Rel. 31, Last sequence update)
DB 20-AUG-2001 (Rel. 40, Last annotation update)
```


QY 316 QPRLKRYKRAMVYOCPPVPOVVOOCTAVOTAAQAMVAFVGVYSSSLPMLSSPSPGO 375
 Db 1547 QO-----OQOQOQOHHOQOQOHHOQANKLLIIPKISMP---SPDKOO 1592
 QY 376 VQTPQSMPPPPQ-PSPQPGQPSQPSNVSSGPAFSPSSFLPSPSPQSPSP---VTAR 430
 Db 1593 LOOHOKVLPQOQSPSDEKILHPNMAAAVAQAQKLVQTKANQONLKNASWSSLSASAN 1652
 QY 431 TPONFVSVPSPPLNTPVNPSVMSVSPAGSS-----QAEQOYLDKLQSLKTYIEPLRMI 484
 Db 1653 SPSHTSSS-----SSSSKAKPAMDSPQOPRNKAKER---DRLLLEAAEKKNOK 1701
 QY 485 NKIDKNEDRKRLKSL 503
 Db 1702 EAAKEQQRKHHSSSSL 1720

RESULT 7

SNFS_YEAST STANLARD; PRT; 905 AA.
 ID SNFS_YEAST
 AC P18460;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-OCT-1994 (Rel. 3C, Last sequence update)
 DE 20-AUG-2001 (Rel. 4C, Last annotation update)
 DE TRANSCRIPTION REGULATORY PROTEIN SNF5 (SWI/SNF COMPLEX COMPONENT SNF5)
 GN (TRANSCRIPTION FACTOR TYE4).
 DE SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MCY;
 RX MEDLINE=91042489; PubMed=2233708;
 RA Laurent B.C., Treitel M.A., Carlson M.;
 RT "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and
 RT proline-rich transcriptional activator that affects expression of a
 RT broad spectrum of genes.";
 RL MOL. CELL. BIOL. 10:5616-5625(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=94378722; PubMed=8091861;
 RA Holmstrom K., Brant T., Kallesoe T.;
 RT "The sequence of a 32,420 bp segment located on the right arm of
 RT chromosome II from Saccharomyces cerevisiae.";
 RL Yeast 10:547-562(1994).
 CC -1- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
 CC COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
 CC OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
 CC BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
 CC -1- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE SNF5 FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M36482; AAA3562.1;
 DR EMBL; X76053; CAA3562.1;
 DR EMBL; Z36158; CAA85254.1;
 DR PIR; S44551; RGHY5.
 DR PIR; S39145; S39145.
 DR SCD; S0000493; SNF5.
 KW Transcription regulation; Activator; Nuclear protein.

FT DOMAIN 31 270 GLN-RICH.
 FT DOMAIN 72 132 PRO-RICH.
 FT DOMAIN 272 324 PRO-RICH.
 FT DOMAIN 489 588 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 714 882 PRO-RICH.
 FT DOMAIN 755 798 ARG/LYS-RICH (BASIC).
 FT CONFLICT 564 564 E->D (IN REF. 1).
 SQ SEQUENCE 905 AA; 102557 MW; A287BA648BD1A35 CRC64;

Query Match 10.8%; Score 326.5; DB 1; Length 905;
 Best Local Similarity 36.8%; Pred. No. 3.6e-07;
 Matches 113; Conservative 22; Mismatches 59; Indels 113; Gaps 16;

QY 46 KKSQASVSDPMNALQSLTGPAAGAGIGMPRGSGISGMSFGAMGQPM---SLSGQ 102
 Db 64 QQQQQQSTSPRQTHQS-----PPPPPGS-----QPIANQSATST 98
 QY 103 PRPGTGMAPHSM-----AVVSTPTPTQOL-QLOQVAL----- 134
 Db 99 PPPPP---APNHLNPIQGVPLRPIRLPQIADLPATQOVLNKLIRQAIAKNNPV 155
 QY 135 -----QOQOQOQ-----QPOQOQAAALQOQOQOQOQOQOQOQOQOQOQO 175
 Db 156 VNAITVAQO 212
 QY 176 AVYQO 235
 Db 213 HHVQIQO 260
 QY 236 QALEAQPTQO 293
 Db 261 Q-----QO 304
 QY 294 ---POSQT 298
 Db 305 NLPKYQT 311

RESULT 8

IP6L_EIMTE STANDARD; PRT; 255 AA.
 ID IP6L_EIMTE
 AC P15714; EIMTE
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DE 01-FEB-1994 (Rel. 28, Last annotation update)
 DE ANTIGEN LPMC-61 (FRAGMENT).
 OS Eimeria tenella.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
 OC Eimeria.
 OX NCBI_TaxId=5802;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Sporozoite; PubMed=2200963;
 RX MEDLINE=90348718; PubMed=2200963;
 RA Ko C., Smith C.K. II, McDonnell M.;
 RT "Identification and characterization of a target antigen of a
 RT monoclonal antibody directed against Eimeria tenella merozoites.";
 RL Mol. Biochem. Parasitol. 41:53-64(1990).
 CC -1- FUNCTION: UNKNOWN. THE GLN-RICH TANDEM REPEATS MAY BE IMPORTANT
 CC FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN
 CC IMPORTANT IMMUNOGEN.
 CC -1- SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER
 CC POLYPEPTIDES TO FORM THE 80 KDA ANTIGEN.
 CC -1- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE
 CC SPOROGATION OF THE OOCYSTS AND IN THE SPOROZOITES FOLLOWING
 CC EXCystation.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

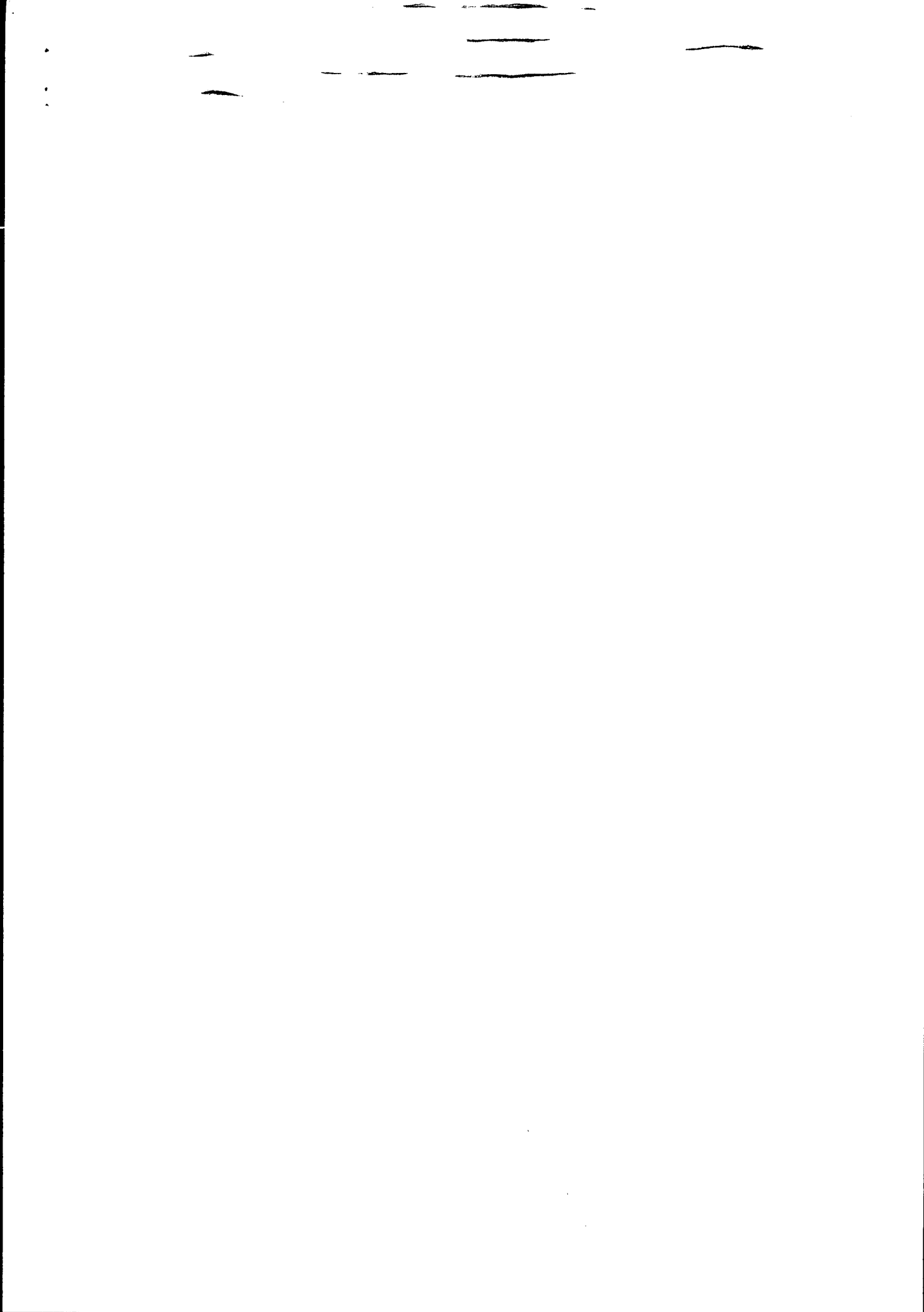
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92069760; PubMed=1720353;
 RA Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
 RT "Prospero is expressed in neuronal precursors and encodes a nuclear
 RT protein that is involved in the control of axonal outgrowth in
 RT Drosophila."
 RT Cell 67:941-953(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92171948; PubMed=1540176;
 RA Matsuzaki F., Koizumi K., Hama C., Yoshioke T., Nabeshima Y.;
 RT "Cloning of the Drosophila prospero gene and its expression in
 RT ganglion mother cells."
 RT Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93083413; PubMed=1842358;
 RA Chu-Iagorai Q., Wright D.M., McNeill L.K., Doe C.Q.;
 RT "The prospero gene encodes a divergent homeodomain protein that
 RT controls neuronal identity in Drosophila."
 RT Development Suppl. 2:79-85(1991).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM PROS-L).
 RX MEDLINE=20503846; PubMed=11051550;
 RA Xu C., Kaufmann R.C., Zhang J., Klady S., Carthew R.W.;
 RT "Overlapping activators and repressors delimit transcriptional
 RT response to receptor tyrosine kinase signals in the Drosophila eye."
 RT Cell 103:87-97(2000).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM PROS-L).
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blake R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Chew J., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklow G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slater E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).
 RN [6]
 RP SIMILARITY TO C. ELEGANS CEH-26.
 RX MEDLINE=94212446; PubMed=7909177;
 RA Buerklin T.R.;
 RT "A Caenorhabditis elegans prospero homologue defines a novel domain."
 RT Trends Biochem. Sci. 19:70-71(1994).
 CC -1- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR
 CC GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS
 CC CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED
 CC FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR
 CC PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT
 CC FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE
 CC TRANSCRIPTION BY BINDING TO DNA.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: PROS-L (SHOWN HERE) AND PROS-
 CC S. ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-
 CC TERMINAL AMINO ACIDS OF THE HOMEODOMAIN.
 CC -1- TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE
 CC DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING
 CC FORMATION.
 CC -1- SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M81389; AAA28841.1; -;
 CC EMBL; D10609; BAA01464.1; -;
 CC EMBL; Z11743; CAA77802.1; -;
 CC EMBL; AF190403; AAF05703.1; -;
 CC EMBL; AE003691; AAF54628.1; ALT. INTR.
 CC PIR; A41089; A41089.
 CC PIR; J01397; J01397.
 CC FLYBASE; FBgn0004595; pros.
 DR Nuclear protein; Transcription regulation; DNA-binding; Homeobox;
 KW Developmental protein; Alternative splicing.
 RN DOMAIN 4 12
 FT DOMAIN 28 31
 FT POLY-ALA.
 FT POLY-SER.
 FT POLY-ASN.
 FT POLY-ALA.
 FT POLY-ALA.
 FT GLN-RICH.
 FT ASN-RICH.
 FT SER-RICH.
 FT POLY-RICH.
 FT POLY-ASP.
 FT POLY-ALA.
 FT GLN-RICH.
 FT POLY-ALA.
 FT POLY-ALA.
 FT NUCLEAR LOCALIZATION SIGNAL.
 FT POLY-PRO.
 FT HIS-RICH.
 FT HOMEBOX (ATYPICAL).
 FT PROSPERO-LIKE.
 FT MISSING (IN ISOFORM PROS-S).
 FT AKMTN1EFGROMKAOATSGLP -> GDAERAWPDEA
 FT GPRGRMFA (IN REF. 1 AND 4).
 FT IGSUNTSKLQOQHNNNSIPANS -> NIALQFHYQVAA
 FT AAATYTLALPIG (IN REF. 1 AND 4).
 FT H -> Q (IN REF. 2).
 FT A -> C (IN REF. 1).
 FT A -> S (IN REF. 1).
 FT T -> S (IN REF. 1).
 FT Q -> QQQQQ (IN REF. 1).
 FT Q -> QQQQQ (IN REF. 1).
 FT SEQUENCE 1403 AA; 153569 MW; 9EF9973E24E238E CRC64;
 SQ
 Query Match 9.9%; Score 298; DB 1; Length 1403;
 Best Local Similarity 22.9%; Pred. No. 7.4e-06;

· Fri Mar 1 09:12:03 2002

us-09-668-119-3.rsp

Page 13




```

Db 181 QQQQQQQQQQQQHL:KIHONQOQIQQQQQQIQRIQALQIQQQQQQQQQQQQALQA 240
QY 241 QPPIQOPPMQOPR'PSCQALPQOLQONHTONHPPOPOOPVAONQPSQLPPOQOTOP 300
Db 241 QPPIQOPPMQOPR'PSCQALPQOLQONHTONHPPOPOOPVAONQPSQLPPOQOTOP 300
QY 301 LVSOQALPQOMLY'QPLK'FVARVAVVO'RVQOVQOQOTAVQTAQAAMVAPGVQVS 360
Db 301 LVSOQALPQOMLY'QPLK'FVARVAVVO'RVQOVQOQOTAVQTAQAAMVAPGVQVS 360
QY 361 QSSLPMLSPSPQCVQOTPOSMPPPOPPSPQOPSPSPSPSPSPSPSPSPSPSPSPSPSP 420
Db 361 QSSLPMLSPSPQCVQOTPOSMPPPOPPSPQOPSPSPSPSPSPSPSPSPSPSPSPSPSP 420
QY 421 QPSQSPVARTPQONE SVSPSPGPIINTFVNPSVSPAGSSQAEQOYLDKLKOLSKYTEPL 480
Db 421 QPSQSPVARTPQONE SVSPSPGPIINTFVNPSVSPAGSSQAEQOYLDKLKOLSKYTEPL 480
QY 481 RRMINKIDKNEDRRK'YDLKSKMKSLDILITDPSKRCPLKTIOKCIATLEKLNDRGCLPHR 540
Db 481 RRMINKIDKNEDRRK'YDLKSKMKSLDILITDPSKRCPLKTIOKCIATLEKLNDRGCLPHR 540
QY 541 PRCHRPMSSTYASR'WMPSPMTSAHLSSTPTCAHSFOP 579
Db 541 PRCHRPMSSTYASR'WMPSPMTSAHLSSTPTCAHSFOP 579

```

```

RESULT 2
ID 015413 PRELIMINARY: PRT: 349 AA.
AC 015413;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 14, Last annotation update)
DE CTG7A (FRAGMENT).
GN CTG7A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBRAL CORTEX;
RC MEDLINE=97369492; PubMed=9225980;
RA Marcolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA Breschel T.S., Stine O.C., Callahan C., McInnis M.C., Ross C.A.,
RA "CDNAS with long CAG trinucleotide repeats from human brain.";
RT Hum. Genet. 100:114-122(1997).
DR EMBL: U80745; AAB91443.1; -.
FT NON-TER 1
FT 349
SQ SEQUENCE 349 AA: 38853 MW: D073F67E22B70533 CRC64:

```

```

Query Match 59.1%; Score 1780; DB 4; Length 349;
Best Local Similarity 99.7%; Pred. No.1.8e-124;
Matches 344; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 161 QQFOAQASAMQOQFQVAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 220
Db 161 QQFOAQASAMQOQFQVAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 220
QY 3 QOFOAQOQXMQOQFQVAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 62
Db 3 QOFOAQOQXMQOQFQVAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 62
QY 221 QQQQQQQQQQQQQQQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 280
Db 221 QQQQQQQQQQQQQQQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 280
QY 63 QQQQQQQQQQQQQQQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 122
Db 63 QQQQQQQQQQQQQQQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 122
QY 281 QPVAQONQPSQLPPO'QOTOP'LVSOQALPQOMLY'QPLK'FVARVAVVO'RVQOVQOQ 340
Db 281 QPVAQONQPSQLPPO'QOTOP'LVSOQALPQOMLY'QPLK'FVARVAVVO'RVQOVQOQ 340
QY 123 QPVAQONQPSQLPPO'QOTOP'LVSOQALPQOMLY'QPLK'FVARVAVVO'RVQOVQOQ 182
Db 123 QPVAQONQPSQLPPO'QOTOP'LVSOQALPQOMLY'QPLK'FVARVAVVO'RVQOVQOQ 182
QY 341 QTAQVTAQAAMVAP'VAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 400
Db 341 QTAQVTAQAAMVAP'VAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 400
QY 183 QTAQVTAQAAMVAP'VAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 242
Db 183 QTAQVTAQAAMVAP'VAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 242

```

```

QY 401 SNVSSGAPAPSPSPSPSPSPSPSPSPVARTPQONE SVSPSPGPIINTFVNPSVSPAGSSQ 460
Db 243 SNVSSGAPAPSPSPSPSPSPSPSPSPVARTPQONE SVSPSPGPIINTFVNPSVSPAGSSQ 302
QY 461 AEEQOYLDKLKOLSKYTEPLRRMINKIDKNEDRRKDKSKMSLTD 505
Db 303 AEEQOYLDKLKOLSKYTEPLRRMINKIDKNEDRRKDKSKMSLTD 347

```

```

RESULT 3
ID 09Y149 PRELIMINARY: PRT: 749 AA.
AC 09Y149;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BCDNA:GH03922 PROTEIN.
GN ABEQOYLDKLKOLSKYTEPLRRMINKIDKNEDRRKDKSKMSLTD 505
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Suter R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Teang G.,
RA Abdayani A., Arcaina T.T., Baxter E., Blazek R.G., Buehner C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacle J.M.,

```



```

QY 372 PGGVOTPSMPPPPSPQPGPS--QPNNSVSGAPSPSSFLPSPSPQSPVTA 429
DB 609 PPMVATVTRGPPPT-----RPAVIAVSSSV--PKQPR---PPTTPPTTRATIR 657
QY 430 RPPONSVPSP-PGLPLTPVNPSSVMPASSQAEBQ-----OYLDKLKOLSKYTEPLR 481
DB 658 PTPONKAVTSRAPAT-----SSVTPSPQSILQOIKRLPAVLILDSKNSTKTEITLL 712
QY 482 R-----MINKIDKNEDRRKKDLSKM 500
DB 713 RDTYGLPLTFYVDKTDKPAIORLOQL 741

RESULT 7
ID 014687 PRELIMINARY; PRT; 4957 AA.
AC 014687;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9738474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560(1997).
DR EMBL: AF010404; AAC51735.1; -.
DR InterPro: IPR000910; HMG_12_box.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001822; Recombinase.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR002965; P-rich_extensn.
DE InterPro: IPR003616; PostSET.
DE InterPro: IPR003888; FYrich_N.
DE InterPro: IPR003889; FYrich_C.
DE Pfam: PF00628; PHD; 3.
DE Pfam: PF00856; SET; 1.
DE PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS00398; RECOMBINASE_2; UNKNOWN_1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00398; HMG; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
SO SEQUENCE 4957 AA; 531848 MW; 1026562E1419CEBD CRC64;

```

```

DB 3327 QQQQQQLQQQQQ---QQQLQQQQQLQQQQQLQQQQQLQQQQQLQQQQQLQQQQQL 3379
QY 217 QQLQLQQQQQQQQQQQ-----QQQQQALHAPPIQQPQQ----- 250
DB 3380 QQQQQQQQ00000000MGLINOSRTLLSPQQ000000VALGPMAPKPLQHPSSGALPTL 3439
QY 251 ----- 250
DB 3440 LITGEQNTVPNAVSSATEGSTHGGSLAIGTPESMATEPEVKPRLSGDSQLLVQ 3499
QY 251 -QPPPPSQ-----ALPQQLQQ---MH----- 268
DB 3500 PGPQPPSSQLQPLRLRPGQQQQQVSLHTAGGSHGQLGSSSSSEASSVPHLLAQSV 3559
QY 269 -----HTQNHOPPPQQPVAQONQPSQTOPPLVSAQALPQOMLYTPPLK 320
DB 3560 SLGDQPGSMTQNLGPPQPMLEPRMNTTGPQP--KGPVLAGGGGLPG--VGIMPTVG 3615
QY 321 FYRAPH--VGOQP-----VOPVOQQ-----GTAVQTAQAQAV-----APGVQS--QS 362
DB 3616 QLRALQGVLANPQLRHLSPQ000QLQALLMQRLQDSQAVRQTPRQEPGTGTFPLQG 3675
QY 363 SL---PMLSS--PSP-----GQVOTPSMPPPP-----QSPQPGQPS 396
DB 3676 LLGCPQLGGFPGPQGTPLQELGAGRPQGPRLPAPRGALSTGVLGVPVHTP-----PP 3731
QY 397 SOPNSVSSGAPSPSSFLPSPSPQSPSPVTAATKPNQFVSVP-----QSPQPGQPS 440
DB 3732 SSPQPKRPSQLPSPSSQLPREALPRTNTPRK-PQGPTEPPRGVSPAAQLADTLF 3790
QY 441 ---GRLNTPVNPSSVMPASS-----QAEEQYLDKOLSKYTEP 479
DB 3791 SKGLGPMPDPDNLATQKPESSSLVPGHLDVNGVYVEASQLSTKQEP 3839

RESULT 8
ID 014686 PRELIMINARY; PRT; 5262 AA.
AC 014686;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9738474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560(1997).
DR EMBL: AF010403; AAC51734.1; -.
DR InterPro: IPR000910; HMG_12_box.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001822; Recombinase.
DR InterPro: IPR001841; Znf_ring.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR003888; FYrich_N.
DR InterPro: IPR003889; FYrich_C.
DR Pfam: PF00628; PHD; 5.
DR Pfam: PF00856; SET; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS00398; RECOMBINASE_2; UNKNOWN_1.

```


Fri Mar 1 09:12:04 2002

us-09-668-119-3.rsp

Page 10

DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00788; RA; 2.
DR ProDom; PD003376; DIL; 1.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00314; RA; 2.
DR ProSite; PS50106; PDZ; 1
Q0 SEQUENCE 1893 AA; 21086 MW; E95BE3470783CDPA CRC64;

Query Match	13.3%;	Score 399;	DB 5;	Length 1893;
Best Local Similarity	27.4%;	Pred. No. 2.4e-21;		
Matches 182; Conservative	58;	Mismatches 196;	Indels 228;	Gaps 30.0;

[illegible]

RESULT	15	
OSW3G1		
ID	OSW3G1	PRELIMINARY;
AC	OSW3G1	PRT; 926 AA.
DT	01-MAY-2000	(REMBLprl. 13, Created)
DT	01-MAY-2000	(REMBLprl. 13, last sequence update)
DT	01-JUN-2001	(REMBLprl. 17, last annotation update)
DE	CG10555	PROTEIN.
GN	CG10555.	
OS	Drosophila melanogaster	(fruit fly).

0C Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
0C Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
0C Ephydroidea; Drosophilidae; Drosophila.
0X NCBI_TaxID=7227;

RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN-BERKELEY;

RA MEDLINE=20196006; PubMed=10731133;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George A.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.-I., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson, Y., Benos P.V., Berman B.F., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cayley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Douf L.E., Downes M., Dugan-Rocha S., Dunkof B.C., Dunn P.

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.
RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacible J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue R.C., Sinden-Kiamos T., Stimson M., Stunski M.P., Smith T.

RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrlskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wand Z.-Y., Wasserman D. A., Weinstein G. M., Weissbach I.

RA Wang Z.-L., Muscatelli L.R., Weir D.J., Yang S.-C., Yao Q.A.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Wang S.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q.,
Zheng L., Zhou Y.-H., Zhou Y.-J., Zhu R.-C., Zhu X.-H., Zhu Y.-C.,
Zhu Y.-H., Zou Y.-R.

[illegible]

KL Science 287:2185-2193(2000).
 DR EMBL: AE003443; AAF46366.1; -
 DR FlyBase: FBgn0030034; CG10555.

```
DR      interpro: IPR002965; P_fich_extensn.
DR      PRINTS: PRO1217; PRICHEXTENS.
SQ      SEQUENCE      926 AA;  93004 MW;  6FD5B9F77C36C006 CRC64;
```

Query Match 13.28; Score 398; DB 5; Length 926;

Best Local Similarity 30.7%; Pred. No. 1.4e-21;
Matches 154; Conservative 24; Mismatches 194; Indels 130; Gaps

QY	50	ASVSDPMNALQSLTGGP-----AAGAAGIGMPRPGRQSLGAGMSFGAMGPMNSLSGQ-	102
Db	189	AQLADPAMNLSQILPPPHILIQAMQGGQQTPTPTSGMLG-----PPQQQQQPPGGG	242

```
QY      103 PPGTSGMADHSMAY-----STATPQTQLQOVALQOQQOQQOQQEQQ   145
```

Db	243	PVFGPGGPFPMGMQHGSDPQGPPYQMPRYGAQQQPQHPLGLPGAQQQSQQQQQQQ	302
Ov	146	000AAI00000000000FOA00SAM000FOAVV00000LO-----0000000HILKTLHH	199

D_b

303 000--000000000000AAAAAAVAGG00GPQVSQGAP0000000HPVV RNA 359

```

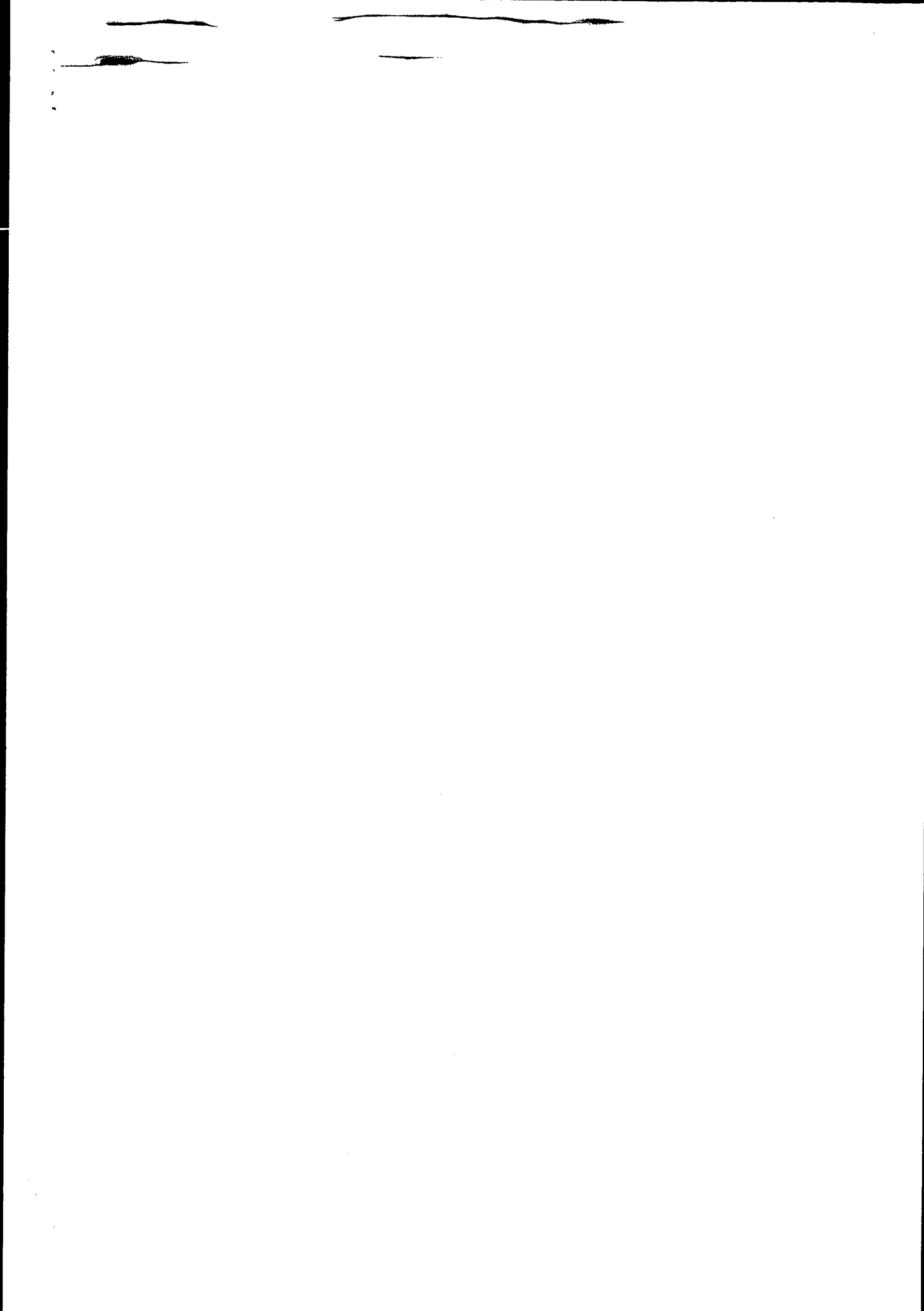
QY 200 QNO-----QOIQQQQQQQLOIRALQLQLQQQQ-----QQQQQQQQQQQALEAQP 243
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 350 QCCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOC 410
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
DU      300  GGGGFGAGGV FGGGGCGVSY INFNKAF NQQRKNNGS LSSGFUNF VVVVVVVVVVF 41
```

Downloaded from ascelibrary.org by University of California, San Diego on 06/01/15. Copyright ASCE, For All Rights Reserved, No part of this document may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without permission in writing from ASCE.

Search completed: February 28, 2002, 19:39:50
Job time: 202 sec



102 AGTTCACAGGCTCAGCAGAGTGCATGTCAGCAGAGATTCCAAAGCAGTAGTG 151

179 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 195

152 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 201

195 eLysLeuHisHisGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 212

202 TAAATTCATCATCAAAATCAACACATACAGCAGCAGCAGCAGCAGCAGCAG 251

212 euglnarGTLleAGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 228

252 TGCACAGCAATAGCAGCAGCTGCAGCTCCAAACAGCAGCAGCAGCAGCAG 301

229 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 245

302 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 351

245 nglnProMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 262

352 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 401

262 lnglnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 278

402 AGCAGCTGACAGAGATGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 451

279 ProGlnGlnProProValAlAGlnAsnGlnProSerGlnLeuProProGln 295

452 CCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 501

295 nSerGlnThrGlnProLeuValSerGlnAlAGlnAlAGlnAlAGlnProGln 312

502 GTCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 551

312 eLeuLeuThrGlnProProLeuLysPheValArgAlaProMetValVal 328

552 TGTGTATATACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 601

329 GlnGlnProProValAlAGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 345

602 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 651

345 nThrAlAGlnAlAGlnAlAGlnAlAGlnAlAGlnAlAGlnAlAGlnAlAGln 362

652 GACAGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 701

362 eSerLeuProMetLeuSerSerProSerProGlnGlnGlnGlnGlnGlnGln 378

702 GCAGTCTCCCATGCTGCTCTGCGGTCACCGGCGCAGCAGCAGCAGCAGCAG 750

379 ProGln...SerMetProProPro 385

751 CCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 774

seq_name: gb_ests2:BG766214

seq_documentation_block: 730 bp mRNA EST 15-MAY-2001

LOCUS BG766214 602737942F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862938 5',

DEFINITION mRNA sequence.

ACCESSION BG766214

VERSION BG766214.1 GI:14076867

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 730)

AUTHOR NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC/DCMP/DRP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>

Plate: L10M1724 row: d column: 11

High quality sequence stop: 716.

Location/Qualifiers

1..730

source

organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4862938"

/clone_lib="NIH_MGC_49"

/tissue_type="melanotic melanoma, high MDR (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 174 a 280 c 172 g 104 t

ORIGIN

alignment_scores:

Quality: 1206.00 Length: 243

Ratio: 5.046 Gaps: 2

Percent Similarity: 98.354 Percent Identity: 97.531

alignment_block:

US-09-668-119-3 x BG766214 ..

Align seg 1/1 to: BG766214 from: 1 to: 730

269 HisThrGlnHisHisGlnProProProGlnProGlnGlnProProValAl 285

4 CACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGC 53

285 aglnAsnGlnProSerGlnLeuProProGlnSerGlnThrGlnProLeu 302

54 TCAAGACCAACATCACACTCCGCGCAGCTGAGCAGCAGCAGCAGCAGCAGTTCG 103

302 alSerGlnAlAGlnAlAGlnProGlnGlnMetLeuThrGlnProPro 318

104 TGTTCACAGCGCGCAAGCTCTCCGAGCAAAATGTTATACCCCAACACCA 153

319 LeuLysPheValArgAlaProMetValAlAGlnGlnProProValGlnPr 335

154 CTGAAATTTGTCGAGCTCGATGATGATGATGATGATGATGATGATGATGATG 203

335 oglnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 352

204 CCAAGTGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 253

352 eValAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 368

254 TGTGTGCTCCCGAGATCCAGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCC 303

369 SerProSerProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 385

304 TCGCGCGCAGCGGCGCAGCAGGTCAGAGACCCCGCACTCATGACCCCTCC 353

385 oProGlnProSerProGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 402

354 CCCCAGGCGCTCCCGCAGCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGTCA 403

402 snValSerSerGlyProAlaProSerProSerSerPheLeuProSerPro 418

/organism="Homo sapiens"

422 ProSerGlnSerProValThrAlaArgThrProGlnAsnPheSerValPr 438

```

seq_name: gb_est2:BG281933
seq_documentation_block:
LOCUS      BG281923          886 bp      mRNA      EST      21-FEB-2001
DEFINITION      602403104rf1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:454543 5',
                mRNA sequence.
ACCESSION      BG281923
VERSION        BG281923.1  GI:13030849
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 886)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robart Strausberg, Ph.D.
               Email: cgaab@remail.nih.gov
               Tissue Procurement: ATCC/DCMP/DMP
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
               Plate: LICM130 row: g column: 02
               High quality sequence stop: 876.
               Location/Qualifiers
                   1
                   886
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_image="454543"
                   /clone_lip="NIH_MGC_20"
                   /tissue_type="melanotic melanoma"

```

/lab host-"DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTR7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCAAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)." .

```
BASE COUNT      234 a      299 c      231 g      122 t
ORIGIN

alignment_scores:
    Quality: 1136.50      Length: 303
    Ratio: 4.209          Gaps: 9
    Percent Similarity: 89.109      Percent Identity: 82.508

alignment_block:
US-09-668-119-3 x BG281923 ..

Align seg 1/1 to: BG281923 from: 1 to: 886

98 SerLeuSerGIyGlnProProProGIYThrSerGIyMeTaLaProHISse 114
   3 TCTCCTTCAGAGTCAGCGCCCTCCTGGACCTCGGGGATGCCCTCACAG 52
114 rMeTaLaVaAlSeTrThrAlaThrProGInPrHnTrGInLeuGIn 131
   53 CAtGCCTGTGTGTGTACGSCAACTCCACAGACCCTGCAAGTCACAC 102
131 InVaLaLaLeuGInGInGInGInGInGInGInPheGInGInGIn 147
   103 AGGTGGCGCTGCAGCAGCAGCAGCAGCAGCAGCATTCAGCAGCAG 152
148 GlNaLaLaLeuGInGInGInGInGInGInGInGInGInPheL 164
   153 CAAGCGCGCTACAGACGACGACGACGACGACGACCAACACAGCAGTCCA 202
164 naLaGInGInSeRaLmeGInGInGInPheGlNaLaVaAlGInGIn 181
   203 GGCTCAACGAGATGCCATGCAGCAGCATTTCCAGCAGTAGTCACAGC 252
181 InGInGInLeuGInGInGInGInGInGInGInHisLeuIleu 197
   253 AGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCATCAATTAAATTG 302
198 HIsHISGInAsnGInGInGInIleGInGInGInGInGInLeuGlnAr 214
   303 CATCATCAAAAATCAGCAAACAGATACAGCAGCAGCAACAGCAGTGCAG 352
214 gILeAlaGInLeuGInLeuGInGInGInGInGInGInGInGInG 231
   353 AATAAGCACAGCTCAGCTCCAMACAACAGCAACAGCAGCAGCAGCAGC 402
231 InGInGIn...GInGInGInAlaLeu.GluAlaGInProProIleGIn.G 246
   403 AGCAGCAACGAGCAGGAGGAGCTTTGGAGAGCCCAAGCAGCAATTCAGCC 452
246 InPrProMetGInGInProGInProProProSerGInAlaLeuProGIn 262
   453 AGCCACCGAATGACAGCAGCCACAGCCTCCGCTCCGAGCTTGCCCCAG 502
263 GlInLeuGInGInMeTHISHISThngInHISHIGInProProProGInPr 279
   503 CAGTGCAGCAGATGCATCACACACAGCAGCAGCAGCGGA.CCAGAGCC 551
279 oGInGInProProVaLaLaGInasnGInProSerGIn.LeuProProGIn 295
   552 CCACAGACCTCCAGTGTCTCAGAGCAGCAGCATCAACAATCTCCGCCAAG 601
296 SerGInThrGInProLeuValSerGInAlaGInAlaLeuProGLyGInMe 312
   |||||||
```

```

602 TCCGACACCCAGCCTTGGTGTCTCAGAGCGGCGAGCTCTCCCTGGACAAAT 651
312 TLeuYrThrGlnProLeuLysPheValAlaAlaPro.MetValVal 328
|||||
652 GTGTATACCAACCAACATGAATTTGTCGAGCTCTGATGTGTGTG 701
329 GlnGln.ProProVal.GlnProGlnValGlnGlnGlnGlnGlnAlaVal 344
|||||
702 CAGCAGGCGCCGAGTGCAGCAGCAGTGCAGCAGCAGCAGCAGTGA 751
345 GlnThrAlaGlnAlaAlaGlnMetValAlaProGlyValGlnValSerGI 361
|||||
752 CAGACCTGCCTCAGCTCCAGATGTGTGTCCCGGAATCCAGATGATCAC 801
361 nSerSerLeuProMetLeu.....SerSerProSerProGlyGlnG 375
|||||
802 GGAAGCTTGGCCAGTGGGATGCACATAGACCCCGTCCCGCTACAA 851
375 lValGlnThrProGlnSerMetProProProProGlnProSerPro 390
|||||
852 CGCTG.....TGTCCGATCCCTCAAGCTCAATCCT 883

```

seq_name: gb_est1:AL046886

seq_documentation_block:

LOCUS AL046886 757 bp mRNA EST 29-FEB-2000
 DEFINITION DKFZP586E2117.T1 586 (synonym: hutel) Homo sapiens cDNA clone
 DKFZP586E2117 5', mRNA sequence.

ACCESSION AL046886

VERSION AL046886.1 GI:5434945

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 757)

REFERENCE Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

TITLE EST (Koehler, et al.)

JOURNAL Unpublished (1999)

COMMENT Contact: Koehler K

MIPS Am Klopferplatz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

sequenced by BMFZ (Biomedical Research Center at the Charite,

Berlin/Germany) within the cDNA sequencing consortium of the German

Genome Project.

No sl sequence available.

This clone (DKFZP586E2117) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..757

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKFZP586E2117"

/clone_id="586 (synonym: hutel)"

/tissue_type="uterus"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pSport1; Site_1: NotI; Site_2: SalI/MluI"

BASE COUNT 199 a 233 c 208 g 116 t 1 others

ORIGIN

alignment_scores:

Quality: 1129.00 Length: 258

Ratio: 4.552 Gaps: 8

Percent Similarity: 96.124 Percent Identity: 94.961

alignment_block:

US-09-668-119-3 x AL046886

Align seq 1/1 to: AL046886 from: 1 to: 757

```

1 MetatrglyAlaGlyValAlaHisSerLysSerSerLysAspMetGluSe 17
|||||
3 ATGAGAAAGCTGTGTGTCACACAGTAATTCACAGCAAGATATGGAAG 52
17 rHisValPheLeuLysAlaLysThrArgAspGluTrpLeuSerLeuVal 34
|||||
53 CCATGTTTCTCGAAGGCCAACACCCCGGAGCAATACCTTCTCTCGGG 102
34 lArgLeuLeileHisPheArgAspIleHisAsnLysSerGlnAla 50
|||||
103 CCAGGCTCATATTCATTTCCAGACATTCATTAACAGAAATCTCAACT 152
51 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyLysAl 67
|||||
153 TCCGTAGTATCTATGATGATGACTCCAGAGCTGACTGCGGAGCTGC 202
67 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSer 84
|||||
203 TGGGGAGCGCGCTGGAATGGCATGCTCTCGGGGCGGGAGAGTCTC 252
84 euGlyIleMetGlySerPheGlyAlaMetGlyGlnProMetSerLeuSer 100
|||||
253 TGGCGGGATGGGTAGCTTGTGTCATGGAGCAGCAGCAATGTCTCTCA 302
101 GlyIleProProProGlyThrSerGlyMetAlaProHisSerMetAla 117
|||||
303 GGGCAGCGGCTCTCTGGAGCTGGGGATGGCCCTCACAGCATGGCTCT 352
117 lValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAla 133
|||||
353 CGTGTCTACGGCAACTCCACAGCCAGCTGCAGCTCCAGAGGTGGCGG 402
134 LeuGlnGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnGlnAla 150
|||||
403 CTCGACGACGACGACGACGACGACGACGACGACGACGACGACGCGG 452
150 lAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAla 166
|||||
453 CGGTACGACGACGACGACGACGACGACGACGACGACGACGACGCTCA 502
166 ngInSerAlaMetGlnGlnGlnPheGlnAlaValValGlnGlnGlnG 182
|||||
503 GCAGAGTGCATGACGACGACGACGACGACGACGACGACGACGACGAC 552
182 ngInLeuGlnGlnGlnGlnGlnGlnGlnGlnHisLeuLeuHis 199
|||||
553 GCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 602
199 lsgInaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 215
|||||
603 ATCAAAATTCAGCAACAGATACAGCAGCAGCAGCAGCAGCAGCGAAT 652
215 eAlaGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 232
|||||
653 AGCAGAGCTGCAGCTC.....CAACAACAGCAACAGCAGCAGCAGC 696
232 lngInGlnGlnGlnAlaLeuGlnAlaGlnProProIleGlnGlnPro 248
|||||
697 AGCAGCAGCAGCAGCAGCTTTGAGGCCAGCCAGCCAAATTCAGCAGC 746
248 oMetGlnGln 251
|||||
747 GATGACAGCAG 756

```

seq_name: gb_est2:BG386117

seq_documentation_block:

LOCUS BG386117 878 bp mRNA EST 12-MAR-2001
 DEFINITION 602455292F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583495 5',
 mRNA sequence.

Fri Mar 1 09:12:04 2002

us-09-668-119-3.rst

Page 6

ACCESSION BG386117
VERSION BG386117.1 31:13279563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 878)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE NIH-MGC http://mgi.mci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rti.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing: Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: NIH Intramural Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1:07 row: h column: 24
High quality sequence stop: 742.

FEATURES
source Location/Qualifiers

1..878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4583495"
/clone_lib="NIH-MGC_15"
/rname_type="adenocarcinoma cell line"
/lat_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 182 a 305 c 240 g 151 t
ORIGIN

alignment_scores:
Quality: 1103.50 Length: 364
Ratio: 4.294 Gaps:
Percent Similarity: 70.604 Percent Identity: 67.857

alignment_block:
US-09-668-119-3 x BG386117 ..

Align seg 1/1 to: BG386117 from: 1 to: 878

27 AspGluTyrLeuSerLeuValAlaArgLeuIleHisPheArgAsp11 43
|||||
4 GAGCAATACCTTCTCTCGTGGCAGGCTCATTCATTTTCGAGACAT 53
|||||
43 eHisAsnLysSerSerAlaSerValSerAspPheMetAsnAlaLeuG 60
|||||
54 TCATACAGAAATCTCAAGCTTCGCTGATCTATGATGACATCTC 103
|||||
60 InSerLeuThrGlyGlyProAlaAlaGlyAlaGlyIleGlyMetPro 76
|||||
104 AAGAGCTGACTGGCGGACCTCTGCGGAGCGGCTGAGATTTGGCATGCTT 153
|||||
77 ProAlaGlyProGlyGlnSerLeuGlyMetGlySerPheGlyAlaMe 93
|||||
154 CCTGGGGGCGGGGACAGTCTCTGGGCGGAGGATGAGCTTGGTGGCAT 203
|||||
93 GAlaGlnProMetSerLeuSerGlyGlnProProGlyThrSerGly 110
|||||
204 GGGACAGCCAAATGCTCTCTCAGGGCAGCGGCTCTGGAGCTGGGGA 253
|||||
110 eAlaProHisSerMeAlaValAlaSerThrAlaThrProGlnThrGln 126

|||||
254 TGCCCTCCACAGCAGCTGCTGTCTACGGCAATCTCCAGACCCAG 303
|||||
127 LeuGlnLeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnPh 143
|||||
304 CTCAGCTCCAGCAGAGTGGCTGCGCAGCAGCAGCAACAGCAGCT 353
|||||
143 eGlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGln 160
|||||
354 C..... 354
160 InGlnGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnPheGlnAla 176
|||||
354 354
177 ValValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 193
|||||
354 354
193 sLeuIleLysLeuHisHisGlnAsnGlnGlnGlnGlnGlnGln 210
|||||
354 354
210 InGlnLeuGlnArgIleAlaGlnLeuGlnLeuGlnGlnGlnGln 226
|||||
354 354
227 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 243
|||||
355 355
243 oGlnGlnGlnProProMetGlnGlnGlnGlnGlnGlnGlnGlnGln 260
|||||
390 AATTAGCAGCCAGCAGATGACAGCAGCCAGCAGCTCCCTCCAGGCT 439
|||||
260 eProGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 276
|||||
440 TGCCCTCCAGCAGCTGAGCAGATGATCAGCAGCAGCAGCAGCAG 489
|||||
277 ProGlnProGlnGlnProProValAlaGlnAsnGlnProSerGlnLeu 293
|||||
490 CCACAGCCCGCAGAGCTCCAGTTGCTCAGAACCAACCATCAACCTCC 539
|||||
293 oProGlnSerGlnThrGlnProLeuValSerGlnAlaGlnAlaLeuPro 310
|||||
540 GCCACAGTGGCAGAC .CAGCTTTGCTGTCACAGCGCAGGCTCTCC 588
|||||
310 LysGlnMetLeuTyrThrGlnProProLeuLys .PheValAlaArgAlaPro 326
|||||
589 GACAAATGTG .TATACCCAAACCACTGAAATTTTTCGAGCTCCGAT 637
|||||
326 tValValGln .GlnProPro .ValGlnProGlnValGlnGlnGln 342
|||||
638 GGTGTGTCAGGAGACCCCGAGTCCAGCCAGGTGACAGCAGCAGAC 687
|||||
342 tAlaValGlnThr .AlaGlnAlaAlaGlnMetValAlaProGlyValGln 358
|||||
688 AGCAGTACAGCCAGCTCAGGTGCTCCAGATGGTGGCTCCGAGATCCAG 737
|||||
359 ValSerGlnSerSerLeu.....PrometLeuSerSerProse 371
|||||
738 ATGATCACGGGGAGGAGCTTGGCCCGAGGTGGAGATGATAGAACCCGG 787
|||||
371 tProGlyGlnGlnVal...GlnThrProGlnSer 381
|||||
788 TCCGCTTACACAGTGTGTGCTCCCTCCCGTAAAC 821
|||||
seq_name: gb_est1:BE384305
seq_documentation_block: 620 bp mRNA EST 21-JUL-2000
LOCUS BE384305
DEFINITION 601272684F1 NIH-MGC_20 Homo sapiens cDNA clone IMAGE:3613868 5',
mRNA sequence.


```

ACCESSION      BE384305
VERSION        BE384305.1 GI:9329670
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 620)
AUTHORS       NIH-MGC http://mgc.ncl.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cga@ncl.nih.gov
                Tissue Procurement: ATCC/DCTD/DRP
FEATURES       CDNA Library Preparation: Ling Hong/Rubin Laboratory
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
                Plate: LICM275 row: g column: 21
                High quality sequence stop: 620.
FEATURES       Location/Qualifiers
                1..620
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3613868"
                /clone_lib="NIH_MGC_20"
                /tissue_type="melanotic melanoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCAAGGAG(G). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT     79 a 148 c 225 g 168 t
ORIGIN
alignment_scores:
    Quality: 1059.00           Length: 206
    Ratio: 5.141              Gaps: 0
Percent Similarity: 100.000   Percent Identity: 99.515
alignment_block:
US-09-668-119-3 x BE384305/rev ..
Align seg 1/1 to reverse of: BE384305 from: 1 to: 620
172 GAGCAGCGAGCAGCATCTAATTAAATTCATCATCAATAATCAGAACAGA 521
620 CACCACTTCCAGCACTACTTGACGAGCAGCAGCAGCTCACACAGAGA 571
188 nginglninglninglhislleuileylseuuhishisglnasnglnglnlt 205
570 GCAGCAGCAGCAGCAGCATCTAATTAAATTCATCATCAATAATCAGAACAGA 521
205 leclninglninglninglninglneuglnarITlAlaglneuglneugln 221
520 TTACAGCAGCGCAGCAGCAGCACTGCACCGAATGACACAGCTGCAGCTCCA 471
222 Glinglncglncglncglncglncglncglncglncglncglncglncgl 238
470 CAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTT 421
238 luguAlaAgInProProIlleGlInglInProPmoetGInglInProGInProp 255
420 GCAGGGCCGACGCCACAATTAGCAGCCACCGATGACGAGCCACACAGCTTC 371
255 roPsoSerGlnAlaLeuProGlInglInleuGlnInmethIshtIrGln 271

```

seq_name: gb_est2:BG336790

```

seq_documentation_block:
LOCUS       BG336790          939 bp      mRNA                      EST
DEFINITION  602405533F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542993 5',
            mRNA sequence.
ACCESSION   BG336790
VERSION     BG336790.1  GI:13143228
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 939)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: L1CML1224 row: a column: 10
            High quality sequence stop: 715.
            Location/Qualifiers
                1..939
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4542993"
                /clone_lib="NIH_MGC_21"
                /tissue_type="Choriocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: Placenta; Vector: pOT87; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit

```


Align seg 1/1 to: BG468341 from: 1 to: 914

```

402 AsnValSerSerGlyProAlaProSerProSerSerPheLeuProSerPr 418
|||||
4 AACGTGAGCTGTGGCCCTGCCCCATCTCCAGTAGCTTCTCGCCGCCGCC 53
|||||
418 oSerProGlnProSerGlnSerProValThrAlaArgThrProGlnAsp 435
|||||
54 CTCACGCGAGCCCTCCAGAGCCAGTACGGGCGGACCCCAACAAGACT 103
|||||
435 heSerValProSerProGlyProLeuAsnThrProValAsnProSerSer 451
|||||
104 TCAGTGTCCCTCCTACCTGGAGCCTTTAAACACACCTGTGAACCCACTCT 153
|||||
452 ValMetSerProAlaGlySerSerGlnAlaGluGluGlnGlnTyrLeuAs 468
|||||
154 GTCATGAGCCGACGCTGCTCCAGCCAGGCTGAGAGACACAGTACTGGA 203
|||||
468 ptySLeuGlyGlnLeuSerTyrTyrTyrTyrTyrTyrTyrTyrTyr 485
|||||
204 CAAGCTGAAGCAGCTGTGCAAGTACATCGAGCCCTGCGCCGCTATGATCA 253
|||||
485 snlySLeaPlyAsnGluAspArgLysLysAspLeuSerLysMetLys 501
|||||
254 ACAAGATCCAGACAGAAAGCAAGAAAGAAAGAGACTGAGTAAGATGAAG 303
|||||
502 SerLeuLeuAspPheLeuThrAspProSerLysArgCysProLeuLysTh 518
|||||
304 AGCCTCTGTGACATCTCTGACAGACCCCTCGAAGCGGTGTCTCCCTGAAG 353
|||||
518 rleuGlnLysCysGluLeuLeuAlaLeuGluLysLeuLysAsnAspMet 534
|||||
354 CTTCGAAAAGTGTGAGTGTGCTCTGAGAAACTCAAGAAATGACATGGCGG 403
|||||
535 CysProLeuProHisArgProArgCysHisArgProAsnSerSerThrTy 551
|||||
404 TGCCCACTCCCGCCACCGCCCGCGTGCAACCGCAACAAGCAGTACCTA 453
|||||
551 rAlaSerArgSerTyrMetProSerTyrProThrSerAlaHisLeuSerS 568
|||||
454 TGGCAGCGCGTCTCTGTGATGCGCTCTGCGCCACATCCGCTGCTGTCTT 503
|||||
568 erThrIleProCysThrAlaHisSerPheGlnPro 579
|||||
504 CAACCATTCCTCTACCGCAGCATTCCTTCCAGCCA 538

```

seq_name: gb_est2:BG746380

seq_documentation_block: 756 bp mRNA

LOCUS BG746380 EST 15-MAY-2001

DEFINITION 602703648F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857083 5',

mrna sequence.

ACCESSION BG746380

VERSION BG746380

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE NIH-MGC http://mgi.ncl.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI708 row: p column: 12
High quality sequence stop: 756.
Location/Qualifiers

FEATURES
source

```

1. /756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4857083"
/tissue="NIH_MGC_15"
/lab="type="adenocarcinoma cell line"
/note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCAAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 179 a 264 c 186 g 126 t 1 others
ORIGIN

```

alignment_scores:
Quality: 919.00 Length: 176
Ratio: 5.251 Gaps: 1
Percent Similarity: 99.432 Percent Identity: 98.864

alignment_block:
US-09-668-119-3 x BG746380 ..

Align seg 1/1 to: BG746380 from: 1 to: 756

```

405 SerGlyProAlaProSerProSerSerPheLeuProSerProSerProG 421
|||||
8 GCAGGCCAGCCCAATCTCCAGTAGCTTCTGCGCCACCCCTCAACCGCA 57
|||||
421 nProSerGlnSerProValThrAlaArgThrProGlnAsnPheSerValP 438
|||||
58 GCCCTCCAGAGCCAGTACGAGCGCGGAGACCCCAACAAGTCAAGTGTCC 107
|||||
438 roSerProGlyProLeuAsnThrProValAsnProSerSerValMetSer 454
|||||
108 CCTCACCTGAGACCTTTAAACACACCTGTGAACCCGCTGTCTGTAGAC 157
|||||
455 ProAlaGlySerSerGlnAlaGluGlnGlnGlnTyrLeuAspLysLeu 471
|||||
158 CCAGCTGGCTCCAGCAGGCTGAGGAGCAGTACCTGGACACAGCTGAA 207
|||||
471 sGlnLeuSerLysTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 488
|||||
208 GCAGCTGTGAGTACATGACGACCCCTGCGCGCATGATCAACAAGATCG 257
|||||
488 sPlyLysAsnGluAspArgLysLysAspLeuSerLysMetLysSerLeu 504
|||||
258 ACNAGAGACGAAAGCAAGAAAAAGAGACTGAGTAAAGATGAGAGCCTTCTG 307
|||||
505 AspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 521
|||||
308 GACATTTGTACAGACCCCTCGAAGCGGTGTCCCTGGAAGACCTTCAAAA 357
|||||
521 sCysGluLeuAlaLeuGluLysLeuLysAsnAspMet 537
|||||
358 GGTGTGAGATCGCCCTGAGAAACTCAAGATGACATGGGGGTGCCACTTC 407
|||||
538 ProHisArgProArgCysHisArgProAsnSerSerThrTyrAlaSerAr 554
|||||
408 CCCACGCGCCCGCGGTGCGCACCGACCAACACAGTACTTATGACGACGG 457
|||||
554 gSerTyrMetProSerTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 571
|||||
458 CTTCCTGGATGCGCTCTGCGCAACATCCGCTACCTGCTTTCACCATTC 507
|||||
571 roCysThrAlaHisSerPheGlnPro 579

```

Fri Mar 1 09:12:04 2002

us-09-668-119-3.rst

Page 10

508 CCTGTACCGCACATTTGTTCAGCCA 533

seq_name: gb_est1:BE019518

```
seq_documentation_block:
LOCUS      BE019568
```

LOCUS	526 bp	MRNA	EST	06-JUN-2000
DEFINITION	ba83f05.y1 N.H.MGC.21 Homo sapiens cDNA clone IMAGE:2907009 5 similar to TR:015413 O15413 CTGTA ; , mRNA sequence.			
ACCESSION	BE019568			

VERSION	22010000	(I:8279646
KEYWORDS	BE019568.1	
SOURCE	EST.	
	human	

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

1

1992

10

67 laalaglalaalaglylleaglymetproparglyproglylginser 83
 578 ccccgagagagagctg3gatgtgacatgctcctccggggcccgagacagctc 529
 84 leuglyglymetgly3serphaglylametglylprometserleuse 100
 528 ct.ggtggatggtggtcccttgcccttgacacacacacacacacacacac 480
 100 tglylproproparglylthrserylmetalaalprohissermetalav 117
 479 cgggagac 430
 117 alvalserthratalatirproglinthrleuglinleuglinvalala 133
 429 tgggtctac 380
 134 leuglinleuglinleuglinleuglinleuglinleuglinleuglin 148
 379 ttgcagcaaac 330
 149 alaalaleu.....glinglinglinglinglinglinglingling 161
 329 gcagac 280
 161 nglnphglnlnaglnlnaglnlnaglnlnaglnlnaglnlnaglnlnagln 178
 279 gcaggtccac 230
 178 alglnleuglnleuglinleuglinleuglinleuglinleuglinleu 194
 229 tgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 186
 195 lleylsleuhsrlslnaglnleuglinleuglinleuglinleuglinleu 211
 185 attagttgacatcatcgaagacacacacacacacacacacacacacacac 139
 211 nleuglnarlgilealnleuglinleuglinleuglinleuglinleuglin 228
 138 actgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 107
 228 lnglnleuglnleuglinleuglinleuglinleuglinleuglinleu 244
 106 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 57
 245 glnleuproprometgln.glnleuproprometserglnleup 261
 56 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 7
 261 ro 261
 6 cc 5

seq_name: gb_est1:BE279307
 seq_documentation_block: 723 bp mRNA 13-JUL-2000
 LOCUS BE279307 601157640F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3504608 5',
 DEFINITION mRNA sequence.
 ACCESSION BE279307
 VERSION BE279307.1 (I:9154297
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 1 (bases 1 to 723)
 REFERENCE NIH-MGC, <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaabts-r@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 Plate: L10M179 row: o column: 09
 High quality sequence stop: 661.
 Location/Qualifiers
 1. 723
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3504608"
 /clone_1b="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pORF7; Site: 1: XhoI;
 Site: 2: EcoRI; cDNA made by oligo-dT priming.
 directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 177 a 208 c 224 g 114 t
 ORIGIN

alignment_scores: Quality: 881.00 Length: 193
 Ratio: 4.686 Gaps: 2
 Percent Similarity: 97.409 Percent Identity: 96.373

alignment_block: US-09-668-119-3 x BE279307 ..
 Align seg 1/1 to: BE279307 From: 1 to: 723

1 MetArgLysAlaGlyValAlaHisSerLysSerSerLysASPmetGluSe 17
 107 ATGAGGAAAGCTGCTGCTGGCACACATTAATCCAGCAAGATATGAGAG 156
 17 rHisValPheLeuLysAlaLysThrArgAspGlyLeuSerLeuVala 34
 157 CCATGTTTCCGTAAGCCCAAGACCCGGGACCAATACCTTCTCTGCTG 206
 34 lArgLeuIleIleHisPheArgAspIleHisAsnLysSerGlnala 50
 207 CCAAGCTCATATTCATTTTCGAGACATTCATTAACAAGAAATCTCAAGCT 256
 51 SerValSerASPProMetAsnAlaLeuGlnSerLeuThrGlyProAl 67
 257 TCCGTAGTGAATCTATGAAATGACATCCAGAGCCCTGACGGGACCTG 306
 67 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSerL 84
 307 TGGGGAGAGCCGCTGGAATGGCATGCTCTCGGGGCGGAGACAGTCTC 356
 84 eucllyglymetglyserphaglylametglylprometserleuser 100
 357 TGGGGGAGATGGGTAGCTTGTGTGACAGGACAGGACCAATGCTCTTCA 406
 101 GlylGlnProProGlylthrserylmetalaalprohissermetalava 117
 407 GGGGAGCCGCTCTCTGGACCTCGGGGATGGCCCTCACAGCATGGCTGT 456
 117 lValSerThrAlaThrProGlnThrGlnleuglinleuglinvalala 134
 457 CGGTGTACGGCAACTCCACAGACCCAGCTGACGTCACGAGGGGGGC 506
 134 euclnleuglnleuglnleuglnleuglnleuglnleuglnleuglnla 150
 507 TGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 556

Fri Mar 1 09:12:04 2002

us-09-668-119-3.rst

Page 14

Result	No.	Score	Query	Match	length	DB	ID	Description
C	1	174.0	100.0	33.4	9	AF056191	AF056191 Homo sapi	
	2	1703.8	97.9	317.0	9	AF328769	AF328769 Homo sapi	
	3	1070.4	61.5	3295	10	AF382770	AF382770 Mus muscu	
	4	1035.8	59.5	1048	9	HS088075	U80745 Homo sapien	
	5	659	37.9	2106	9	BC007529	BC007529 Homo sapi	
	6	560	32.2	156315	2	AC068137	AC068137 Homo sapi	
	7	353.4	20.3	64231	2	AL359452	AL359452 Homo sapi	
	8	353.4	20.3	157080	2	AL358856	AL358856 Homo sapi	
	9	353.4	20.3	163908	2	AC007050	AC007050 Homo sapi	
	10	340.4	19.6	145356	6	AC004033	AC004033 Homo sapi	
	11	338.6	19.5	386	6	AF071270	AF071270 Sequence	
	12	220.8	12.7	238442	2	AC090437	AC090437 Mus muscu	
	13	217.8	12.5	4701	2	AK000003	AK000003 Homo sapi	
	14	209.8	12.1	121330	2	AC091616	AC091616 Rattus no	
	15	209.6	12.0	187315	2	AC087802	AC087802 Mus muscu	
	16	209.6	12.0	228467	2	AC079044	AC079044 Mus muscu	
	17	207.2	11.9	5719	10	MMCAT51	AF051726 Mus muscu	
	18	206.8	11.9	162869	9	AC090645	AC090645 Homo sapi	
	19	201.2	11.6	168874	33	AC021022	AC021022 Homo sapi	
	20	201.2	11.6	168778	9	AC090886	AC090886 Homo sapi	
	21	201.2	11.6	169590	9	AC090004	AC090004 Homo sapi	
	22	201.2	11.6	209973	2	AC073755	AC073755 Mus muscu	
	23	198	11.4	2479	10	RN087960	U87960 Rattus norve	
	24	197.8	11.4	3941	10	RN1842	X83546 R.norvegicus	
	25	195.8	11.3	198470	2	AC090121	AC090121 Mus muscu	
	26	191.4	11.0	197321	2	AC092203	AC092203 Mus muscu	
	27	188.8	10.9	179553	2	AC024253	AC024253 Homo sapi	
	28	185.4	10.7	197232	2	AC073946	AC073946 Mus muscu	
	29	183.8	10.6	1611	10	MM070651	U70651 Mus musculus	
	30	183.6	10.6	1612	10	MM070652	U70652 Mus musculus	
	31	182.2	10.5	9553	5	AF134321	AF134321 Dissoscic	
	32	181.8	10.4	335884	2	AC092202	AC092202 Mus muscu	
	33	181.6	10.4	1586	10	MM070653	U70653 Mus musculus	
	34	181.6	10.4	1598	10	MM070654	U70654 Mus musculus	
	35	181.2	10.4	3489	6	AR072952	AR072952 Sequence	
	36	181.2	10.4	3489	6	AX107938	AX107938 Sequence	
	37	181.2	10.4	3489	14	KSU52064	U52064 Kaposi's sar	
	38	181.2	10.4	32207	6	AR055852	AR055852 Sequence	
	39	181.2	10.4	32207	6	AR127850	AR127850 Sequence	
	40	181.2	10.4	137508	14	KSU75598	U75598 Kaposi's sar	
	41	180.8	10.4	28559	14	AF148805	AF148805 Kaposi's	
	42	180.6	10.4	186935	2	AC022332	AC022332 Homo sapi	
	43	177.4	10.2	180658	2	AC020857	AC020857 Mus muscu	
	44	176.4	10.1	2070	5	BS043200	U43200 Boreogadus	
	45	174.4	10.0	1968	10	MM046463	U46463 Mus musculus	

Fri Mar 1 09:11:57 2002

us-09-668-119-1.rge

Page 2

FEATURES
SOURCE
JOURNAL
TITLE
Direct Submission
Submitted (33-MAR-1998) Microbiology, State University of New York
Health Science Center at Brooklyn, 450 Clarkson Ave., Brooklyn, NY
11203, USA
location/Qualifiers
1 22324

CDS

TPA inducible gene; TIG-1"

```
/product="TPA inducible protein"  
/protein_id="AAC12944.1"  
/db_xref="GI:3037135"
```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

BASE COUNT	800 a	1088 c	893 g	553 t
ORIGIN				

Query Match	100.0%	Score 1740;	DB 9;	Length 3334;
Best Local Similarity	100.0%	Pred. No. 5.4e-256;		
Matches 1740; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible][illegible]

QY 583 attaatgcatcatcaataatcagcaagatatacagcagcaagcaagcagctgacagcga 642
 Db 673 ATTAAATTGATCATATCAAGCCAGCAACA--ACAGATATACACACACACTGACAGAG 729
 QY 643 atagacagcttgagctcacaacacagcaacagcagcagcagcagcagcagcagcagcag 702
 Db 730 ATGGCAAGATTGACACT-----GCAACAACAGCAACAGAGCAACAG 771
 QY 703 cagcagcgttgaggccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 762
 Db 772 CAGCAGGCTTTTCAGGCCCCAGCCAAATGCGAGCCCATCATCAATCAGCAGCAGCT 831
 QY 763 cagcagcgttgagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 822
 Db 832 CCCCCCTTCTCAGGCCCCCTACCCAGAGCTGTACAGCTGATCATCCACAGCATACAG 891
 QY 823 ccgacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 882
 Db 892 CCACACCTCAAGCTCAGAGAGTCCCAATGCTCAAAACCAACACACAGATGCCACCA 951
 QY 883 cagctgcagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 941
 Db 952 CAGTACACAGAGCAGCTTTGGTGTACAGAGCAAGCCCTTCCAGACCGATGCTGTAT 1011
 QY 942 --taaccaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 999
 Db 1012 GCTGCCACACAGCAGCTGAATTTTCTGCTCCGATGATGTGTCCAGCAGCCGCAAGTG 1071
 QY 1000 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1041
 Db 1072 CAGCCCCAGAGTGCAGAGTGCAGCCCCAGGTCAGAGCCGACAGCAGAGTGCAGAGCAGA 1131
 QY 1042 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1071
 Db 1132 CAGTGTGCCAGATGTAGTCTCCGAGCTCAGATGATGTTGTAAGACCTTGGCCCAAGGC 1191
 QY 1072 ----- 1071
 Db 1192 GGGATCAGAGTAAGAGCCGGTTCCGCCACTCCACCATGTCGTGGCCCGTCAAGC 1251
 QY 1072 -----cagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1101
 Db 1252 TCCATCTTTTGGGCGAGAGCCACACACAGAGTGCAGCAAGCCTCCTACCATGCTG 1311
 QY 1102 tctctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1161
 Db 1312 TCTTCAACCTTCAACAGGCGAGGTCAGACCCACAGTGTGACCTCCCCACAG 1371
 QY 1162 ccgtcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1221
 Db 1372 CGGTCCCAACACTG-----GCTCAGAGCCCACTTAATGTACGTCCGCGCT 1422
 QY 1222 gcccaatctccagtagcttctcagcagcagcagcagcagcagcagcagcagcagcagcagc 1281
 Db 1423 GCCCATCTTCCAGCAGCTCTGCTTACGCTTACACACACCTTCTCAGAGCCGCTG 1482
 QY 1282 aagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1341
 Db 1483 ACAGCAGGAGCCCAAGAGACTTACGCTTCTCCCTGAGACTTAAACACCCCTGTG 1542
 QY 1342 aacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1401
 Db 1543 AACCCAGAGCTGTATGAGCCAGCTGTAGCAGAGCTGAGAGAGCAGATACCTG 1602
 QY 1402 gacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1461
 Db 1603 GACAACTGAAAGCAATTTGCCAAGTACATGAGACCCCTCGAGCATGATCAACAAGATC 1662
 QY 1462 gacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1521
 Db 1663 GACAAAGATTAACACAGAAAAAGACTTAAGTAAGATGAAGAGCCTGCTGACATCCTC 1722
 QY 1522 acagacccctcgaagcgtgtccctgaagaccttgcaaaatgtgagatcgcccttgag 1581

Db 1723 ACGAGCCCTGAGAGAGTGTCCCTGTAAGACCTGTGAAAAAGTGTGATGCTGGAGAG 1782
 QY 1582 aaactcaagaatgacat-gcagtgccacatcccccagccagcagcagcagcagcagcagc 1640
 Db 1783 AAGCTCAAGATGACATGAGGAGTGGCCACACCCCCACACACCCAGTTCTTCCAAACCA 1842
 QY 1641 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1700
 Db 1843 CAGCAGAGCCTGTGCCAACCACTCTGTAGTGAAGTCTTGCCACATCCGTTCACTGTC 1902
 QY 1701 ttcaacatccctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740
 Db 1903 TTCAACCATTCCTGTGACGACATTTGTGACAGCATGA 1942

RESULT 4

HSU80745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

repeat_region

BASE COUNT 269 a 388 c 251 g 139 t 1 others
 ORIGIN /rpt_type=tandem
 /rpt_unit=CAG

Query Match 59.5%; Score 1035.8; DB 9; Length 1048;
 Best local Similarity 99.7%; Pred. No. 1,2e-148;
 Matches 1037; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH080A13
----- Summary Statistics -----
Sequencing vector: M13; 38%
Sequencing vector: plasmid; 22%
Chemistry: Dye-primer ET; 2% of reads
Chemistry: Dye-terminator Big Dye; 58% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 153264 bases at least Q40
Consensus quality: 156105 bases at least Q30
Consensus quality: 157587 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 169020; sum-of-contigs
Quality coverage: 9.85 in Q20 bases; agarose-fp
Quality coverage: 9.86 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
* 1          9172: contig of 9172 bp in length
* *
* 9173       9272: gap of unknown length
* *
* 9273       17587: contig of 8315 bp in length
* *
* 17588      17687: gap of unknown length
* *
* 17688      39085: contig of 21398 bp in length
* *
* 39086      39185: gap of unknown length
* *
* 39186      61924: contig of 22739 bp in length
* *
* 61925      62024: gap of unknown length
* *
* 62025      107416: contig of 45392 bp in length
* *
* 107417     107516: gap of unknown length
* *
* 107517     156315: contig of 48799 bp in length.
*
* Location/Qualifiers
*   1..156315
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="2"
*     /clone="RP11-803A13"
*       1..9172
*         /note="assembly_name:Contig66"
*           9273..17587
*             /note="assembly_name:Contig67"
*               17688..39085
*                 /note="assembly_name:Contig68"
*                   39186..61924
*                     /note="assembly_name:Contig69
*                       clone_end:SP6
*                         vector_side:right"
*                           62025..107416
*                             /note="assembly_name:Contig70"
*                               107517..156315
*                                 /note="assembly_name:Contig71
*                                   clone_end:T7
*                                     vector_side:right"
BASE COUNT    42538 a 35745 c 36389 g 41139 t    504 others
ORIGIN
Query Match            32.2%; Score 560; DB 2; Length 156315;
Best Local Similarity 77.0%; Pred.No.1.4e-76;
Matches 865; Conservative 0; Mismatches 210; Indels 49; Gaps 13
QY      648 acagctgcagctcaacaacagcaaacagcagcagcgacgacgacgacgacga 707
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 154393 AAACACCACAACGAGCAGCAGCAAGGCCAACAGCTGCAGCTACGAGCAGCAGCA 154334
QY      708 ggatttgaggccagccaccacaaatcaagcagcacgatga--gcagccaagcttc 764
        ||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154333 GGCTTTGCGAGGCCCCAGTG -CCATTAAGCAGCGCCACCACCATGTGCATTTCGTGCGAGCTTGC 154275

```


FEATURES	
*	88143 88242: gap of 100 bp
*	88243 95033: contig of 6791 bp in length
*	95034 95133: gap of 100 bp
*	95134 100091: contig of 4958 bp in length
*	100092 100191: gap of 100 bp
*	100192 117326: contig of 17135 bp in length
*	117327 117426: gap of 100 bp
*	117427 115946: contig of 2520 bp in length
*	115947 120046: gap of 100 bp
*	120047 130509: contig of 1463 bp in length
*	130510 130609: gap of 100 bp
*	130610 139309: contig of 8700 bp in length
*	139310 139409: gap of 100 bp
*	139410 157080: contig of 17671 bp in length
	Location/Qualifiers

```

misc_feature      37302 a 33841 c 39467 g 39369 t 1101 others
/misc_feature     /no.e="assembly-fragment:00230
/misc_feature     clo.e_end:SP6
/misc_feature     vec.eor.side:left"
/misc_feature     3561.13200
/misc_feature     /no.e="assembly-fragment:00057
/misc_feature     fragment_chain:1"
/misc_feature     13301.39189
/misc_feature     /no.e="assembly-fragment:01204
/misc_feature     fragment_chain:1"
/misc_feature     59210.85824
/misc_feature     /no.e="assembly-fragment:01664
/misc_feature     fragment_chain:1"
/misc_feature     85935.88142
/misc_feature     /no.e="assembly-fragment:00505
/misc_feature     fragment_chain:1"
/misc_feature     88213.95033
/misc_feature     /no.e="assembly-fragment:01296
/misc_feature     fragment_chain:2"
/misc_feature     95114.100091
/misc_feature     /no.e="assembly-fragment:01408
/misc_feature     fragment_chain:2"
/misc_feature     100.92.117326
/misc_feature     /no.e="assembly-fragment:00869
/misc_feature     fragment_chain:2"
/misc_feature     117127.119946
/misc_feature     /no.e="assembly-fragment:00628
/misc_feature     fragment_chain:3"
/misc_feature     120047.130509
/misc_feature     /no.e="assembly-fragment:01620
/misc_feature     fragment_chain:3"
/misc_feature     130610.139309
/misc_feature     /no.e="assembly-fragment:01199"
/misc_feature     13910.157060
/misc_feature     /no.e="assembly-fragment:01779"
BASE COUNT      37302 a 33841 c 39467 g 39369 t 1101 others
ORIGIN

```

Query Match	20.3%;	Score 353.4;	DB 2;	Length 157080;
Best Local Similarity	99.7%;	Pred. No. 4.2e-45;		
Matches 354;	Conservative	0;	Mismatches 1;	Indels 0;
			Gaps	0

[illegible]

Db	75386	TTGAGCAGCCAGATGACGAGCCACAGCCCTCCGCCCTCCAGAGCTTCGCCACAGAC	75327
Qy	791	tgcagcagatgcatgacacacagacacacacgcccacacagcccgagcctccag	850
Db	75326	TGCAGCAGATGCATTCACACACGACACACACGCGCCACACACGCCACGACCTCCAG	75267
Qy	851	tctgctagaacacacatcacactccgcgcacagtcgcagaccagcctttgltgac	910
Db	75266	TTGCTCAGAACCCACACATACACACTCCGCCACAGTCGAGACCCAGCCTTGGTGTAC	75207
Qy	911	aggcgcaagctccctctggacaacatgttgttaccacaacacacgaatttgc	965
Db	75206	AGGCGCAAGTCTCTCCCTGACAAATGTTGTATTACCAACACACACTGAATTTGT	75152

RESULT	9
AC007050	
LOCUS	AC007050 163908 bp DNA PRI 31-MAY-2000
DEFINITION	Homo sapiens chromosome 22q11 clone bac32, complete sequence.
ACCESSION	AC007050
VERSION	AC007050.25 GI:6456174
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL	(Pages 1 to 163908) Chen,F., Hu,X., Budarf,M. and Roe,B.A. Homo sapiens Chromosome 22q11 BAC Clone bac32 In BCRL2-GGT Region Unpublished 2 (Pages 1 to 163908) Chen,F., Hu,X., Emanuel,B., Budarf,M. and Roe,B.A. Direct Submission Submitted (09-MAR-1999) Department Of Chemistry And Biochemistry,
--	--

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 163908)
Chen, F., Hu, X., Budarf, M., Emanuel, B. and Roe, B.A.
Direct Submission
Submitted (07-JUL-1999) Department Of Chemistry And Biochemistry,

REFERENCE
AUTHORS
TITLE
JOURNAL

4 (pages 1 to 161908)
Chen, F., Hu, X., Budarf, M., Emanuel, B. and Roe, B.A.
Direct Submission
Submitted (19-JUL-1999) Department Of Chemistry And Biochemistry,

REFERENCE
5 (Pages 1 to 163908)
AUTHORS Chen, F., Hu, X., Budarf, M., and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1999) Department of Chemistry And Biochemistry,
The University of California, San Diego, La Jolla, California 92037, U.S.A.

REFERENCE
AUTHORS
TITLE
JOURNAL

6 (bases 1 to 163908)
Chen, F., Hu, X., Budarf, M., Emanuel, B. and Roe, B.A.
Direct Submission
Submitted (29-JUL-1999) Department of Chemistry And Biochemistry,

REFERENCE
AUTHORS
TITLE
JOURNAL

7 (bases 1 to 163908)
Chen, F., Hu, X., Budarf, M., Emanuel, B. and Roe, B.A.
Direct Submission
Submitted (26-AUG-1999) Department of Chemistry And Biochemistry,

REFERENCE 8 (bases 1 to 163908)
AUTHORS Chen, F., Hu, X., Budarf, M., Emanuel, B. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE	9 (bases 1 to 163908)
AUTHORS	Chen,F., Hu,X., Budarf,M., Emmanuel,B. and Roe,B.A.
TITLE	Direct Submission
JOURNAL	Submitted (19-Oct-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE	10 (bases 1 to 163908)
AUTHORS	Chen,F., Hu,X., Budarf,M., Emmanuel,B. and Roe,B.A.
TITLE	Direct Submission
JOURNAL	Submitted (20-Nov-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE	11 (bases 1 to 163908)
AUTHORS	Chen,F., Hu,X., Budarf,M., Emmanuel,B. and Roe,B.A.
TITLE	Direct Submission
JOURNAL	Submitted (31-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT	On Nov 20, 1999 this sequence version replaced gi:6067157. Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below.
FEATURES	AC004033(bp_m1) 89227..145356 (0) overlaps AC007050(bac32) 1 59304 (104604) AC007050(bac32) 141234..163908 (0) overlaps AC007308(pac408) 1 22675..(193113). Location/Qualifiers 1..163908 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="2q11" /clone="bac32"
BASE COUNT	39327 a 41642 c 42088 g 40851 t
ORIGIN	
Query Match	20.3%, Score 353.4; DB 9; Length 163908;
Best Local Similarity	99.7%; Pred. No. 4.2e-45;
Matches 354; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Db 611	agatcagcagcagaacacagcagcgtgcgcgaatgacagctgcagctccaacacagc 670
Db 14559	AGATCAGCAGCAGCAACACAGCGCTGCAGCGCAATAGCAGCTGCACACACACAGC 14618
QY 671	aacagcagcagcagcagcagcagcagcagcagcagcagcgtttgagggccacacaa 730
Db 14619	AACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTGCAGGCGCCACACAA 14678
QY 731	ttcagcagcagcagcgttgagcagcagcagcagcgtccgcctccagagctctgcgccagcagc 790
Db 14679	TTTCAGCAGCGCCGATGAGCGACGCCACACAGCTTCGCGCTCCAGGCTCTGGCCCCAGCAGC 14738
QY 791	tgcagcagaatgatcacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 850
Db 14739	TGCACGAGATGATCATCACACAGCACACACACAGCGCCGACACAGCGCCGACAGCGCTCCAG 14798
QY 851	ttgctcagaacacacatcacatccctccggccacagctcgagagccagcctttggtgtcac 910
Db 14799	TTGCTCAAGAACCAACCATATCACAACATCCGCCGACAGCTGCAGACCAAGCCTTTGGGTGTCA 14858
QY 911	aggcgcaagctctccctgagacaatggtgtataccaacaacacacactgaattgt 965
Db 14859	AGGCGCAAGCTCTCCCTGSGAATAATTGTGTATACCCAACACACACAGCAATTTGT 14913
RESULT 10	
AC004033	145356 bp DNA PRI 31-MAY-2000
LOCUS	Homo sapiens Chromosome 22q11.2 PAC Clone P_m11 In BCR12-CGT
DEFINITION	Region, complete sequence.
ACCESSION	AC004033
VERSION	AC004033.3 GI:4581183
KEYWORDS	HTG.
SOURCE	human

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE AUTHORS JOURNAL TITLE	1 (bases 1 to 145356) Budarf,M.L. and Emanuel,B.S. Unpublished 2 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Homo sapiens Chromosome 22q11.2 BAC Clone p_m11 In BCRL2-GGT Region Unpublished 3 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (26-JAN-1998) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 4 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (09-DEC-1998) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 5 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (26-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 6 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (04-FEB-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 7 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (09-APR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 8 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (12-APR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 9 (bases 1 to 145356)
REFERENCE AUTHORS JOURNAL TITLE	Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A. Direct Submission Submitted (31-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA On Apr 12, 1999 this sequence version replaced gi:14580479.
COMMENT	Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below. AC00500(p52d6) 112552 192592 (U) overlaps AC004033(P.m11) 1 78509 (p68447) AC004033(P.m11) 89227 145356 (U) overlaps AC007050(dac32) 1 59504 (104601).
FEATURES	Location/Qualifiers source 1..145356 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /chromosome="22q11.2" /map="22q11.2" /clone="p_m11"
BASE COUNT	34453 a 36720 c 37358 g 36825 t
ORIGIN	
Query Match	19.6%; Score 340.4; DB 9; Length 145356;
Best Local Similarity	98.9%; Pred.No. 4.1e-43;

Fri Mar 1 09:11:57 2002

us-09-668-119-1.rge

Page 16

Db 72241 CAGAGCGCCGCTTGTGTGTCACAGCAGCAGCCCTTCTGTGACGATGCTGTATGCTGCC 72182

Qy 946 caaccacactgaatttgt 965
|||||
|||||

Db 72181 CAGAGCGCCTGAAATTGT 72162

Search completed: February 28, 2002, 19:26:56
Job time: 10749 sec